Journal of
Experimental Psychology

VOL. 37, NO. 2 APRIL, 1947

AN ATTEMPT TO CONDITION THE *CHRONIC* SPINAL DOG*

BY W. N. KELLOGG, JAMES DEESE, N. H. PRONKO, AND M. FEINBERG

Conditioning Laboratory, Indiana University

The remarkable conditioning experiments of Culler and Shurrager on acute spinal dogs (**1, 21, 22, 23**) have led to the conclusion that simple learning can take place, not only without the aid of the cerebral cortex, but also without direct neural connections to the basal ganglia, the cerebellum, and the medulla oblongata. The method employed by these investigators was to transect the spinal cord of the experimental animal at some point anterior to the third lumbar roots and at the same time to expose the flexing (semitendinosus) muscles of the hind legs. The response to be conditioned was the contraction or twitching of the distal end of the semitendinosus muscle. In order to eliminate uncontrolled movements on the part of the preparation, it was held in a rigid position upon its back by means of a drill-rod which pierced both of the femurs.

Upon the completion of this surgery, that part of the animal posterior to the cord transection was given conditioned-reflex training. The conditioned stimulus was in most cases three consecutive a.c. electric shocks—spaced one sec. apart—which were delivered to the tail. The unconditioned stimulus was an a.c. electric shock to the hind foot of sufficient intensity to elicit a full contraction of the homolateral semitendinosus muscle. The unconditioned stimulus on any given trial occurred synchronously with the third impulse of the conditioned stimulus.

Before training, the shocks to the tail produced no observable response in the semitendinosus muscle. But after pairing of the conditioned and unconditioned stimuli, the first two tail shocks by

* This investigation was financed by a series of grants-in-aid of research from the Graduate School of Indiana University. The present paper is a follow-up of the preliminary report which appeared in *Science* something over a year ago (**10**).

themselves would then cause the semitendinosus muscle to twitch or contract. This phenomenon was observed in 98 of 219 animals of the original group studied. In some cases it appeared in a surprisingly small number of trials and would reach the level of efficiency of 10 successive conditioned responses in 10 trials, within as few as 30 or even 20 trials of training.

Results like those of Culler and Shurrager on spinal conditioning follow as a kind of logical sequence from the original experiments of Lashley (**13**) which showed, in general, that no special area of the cerebral cortex was prepotent over any other area for the development of simple motor learning. A further step in this direction was the demonstration by Culler and his associates (**2, 3, 4, 15**) and also by Lebedinskaia and Rosenthal (**14**) that conditioning could occur in the dog after complete—or very nearly complete—decortication. The final step of sectioning the cord leaves one with the impression that rudimentary learning may require only the basic sensory and motor fibers which connect the stimulated and responding body parts.

To carry this line of reasoning to its ultimate conclusion, it may be argued that conditioning should be possible in an isolated organ of the body without neural attachments of any kind whatever, providing effective stimulation and a method for measuring the response were available. A phenomenon suggestive of conditioning of this sort has in fact been observed in the isolated heart of the frog by Sgonina (**19, 20**). The excised organ was maintained in a watchglass in Ringer's solution. The original or unconditioned stimulus was a rise in the temperature of the surrounding medium which produced a response of *acceleration* in the heart rate. The conditioned or 'signal' stimulus was a mechanical stroking of the heart, which seems to have caused no reaction at all when first presented by itself. After pairing of the two stimuli, both the conditioned stimulus alone, and the unconditioned stimulus alone, now produced a *retardation* in the heart rate. Since the responses before and after 'training' were different and in fact antagonistic, it would appear that the phenomenon observed by Sgonina was really not conditioning, in the ordinary meaning of that term. It was rather a kind of sensitization (*Sensibilisierung*), or reduction in the threshold for a previously absent reaction.

Problem and Conditions

The implications of the findings on spinal conditioning for physiology, neuropsychology, and the psychology of learning are so far-reaching that they strike at the very roots of the general conception of the nervous system. The present experiments were planned,

therefore, in the attempt to probe further into the study of this extraordinary spinal phenomenon. An effort was made to keep the experimental conditions similar to those of the original investigations, so that the results obtained would serve as a check or corroboration of the earlier studies. At the same time, a number of changes in technique were introduced which, it was hoped, would improve the method.

The most important of these changes was the use of *chronic* rather than *acute* preparations. This necessarily had the effect of greatly reducing the number of cases, when compared to those employed by Culler and Shurrager. But it also permitted the giving of a much larger number of trials per case than would be possible with preparations which were trained and discarded on the day of operation. There were other differences in method which will be discussed where they are relevant to the present results. That these differences may have influenced the data to be reported is an unquestioned possibility. Their importance should, therefore, in no way be minimized.

Experimental Procedure

Control experiments.—It is obviously necessary in spinal conditioning to employ either tactual or kinesthetic stimulation and at the same time to differentiate adequately between the original or unconditioned stimulus, and the conditioned stimulus. In the present experiments this differentiation was accomplished by sending the conditioned stimulus to the left-hind foot and the unconditioned stimulus to the right-hind foot—which therefore served as the member to be conditioned. The stimuli were further differentiated by the fact that the conditioned stimulus was 1000-cycle alternating current while the unconditioned stimulus was a make-break d.c. shock .2 sec. in duration.

Both conditioned and unconditioned stimuli were regulated in intensity in accordance with the magnitude of the reflexes they elicited in the two hind limbs. Their intensity was adjusted so as to give repeated flexion reflexes of about four in. in each of the members to which the stimuli were delivered. The advantage of this procedure is that there was never any doubt as to whether or not a particular stimulus—either conditioned or unconditioned—had been 'received.' The size of each reflex response was *prima facie* evidence of the effectiveness of the stimulus which produced it.

To test out the efficiency of the shock-shock (or double shock) conditioning method when the shocks were delivered to the two hind feet, two control groups of eight intact laboratory dogs were run—four to a group—with different temporal arrangements of the conditioned and unconditioned stimuli. In the first of these groups the shock-shock technique was similar to that employed by Pronko and Kellogg (16, 17). That is, the conditioned stimulus (to the left-hind foot in the present instance) was a continuous a.c. shock of two full sec. duration (see right-hand record, Fig. 2), while the unconditioned stimulus (to the right-hind foot) was the make-break d.c. impulse .2 sec. in duration, which occurred simultaneously with the last .2 sec. of the conditioned stimulus. In the second of these groups the temporal arrangement was similar, but not identical, to that employed by Culler and Shurrager. The conditioned stimuli on every trial were two separate 1000-cycle shocks—each .2 sec. in length and spaced one sec. apart. The unconditioned stimulus was a d.c. shock .2 sec. long which came one sec. *after* the last of the a.c. shocks (see left-hand record, Fig. 2). The fundamental difference between this sequence of stimuli and the Culler and Shurrager technique is that the former investigators presented three a.c. shocks as conditioned stimuli on each trial. The third conditioned impulse (to the tail) coincided with the unconditioned shock to the foot. Since our own method never presented two simultaneous shocks, but eliminated the final conditioned stimulus employed by Culler and Shurrager, it would be classified as a definite instance of forward conditioning.

The effectiveness of the method using a long-shock conditioned stimulus—when the shocks were delivered to the two hind feet in normal dogs—was found by these control experiments to be about equivalent to the effectiveness of this method when the conditioned stimulus was sent to the diagonal forelimb (17). Moreover, the short double-shock conditioned stimulus, when the shocks were sent to the hind feet, gave only slightly inferior results to the long-shock method. The difference was so small that it could be ascribed to individual variation in the *S*s of the two groups.

As a result of these findings, the second of the two stimulating patterns was adopted for spinal conditioning in the present study. The stimulus sequence employed was therefore as follows.

(a) A conditioned stimulus consisting of two 2-sec. a.c. impulses to the left-hind foot, spaced one sec. apart, followed by

(b) An unconditioned stimulus consisting of a make-break d.c. impulse, .2 sec. long, delivered to the right-hind foot one sec. later. The conditioned and unconditioned impulses together encompassed a time interval of about two full sec.

The experimental situation.—All *S*s for both control and spinal experiments were unselected laboratory dogs weighing from 20 to 40 lbs. which were purchased from an animal dealer. The conditioning was done in a sound-proofed chamber where the *S* could be observed through one-way-vision windows. The general experimental environment has been described in detail elsewhere (9).

The spinal preparations were not inverted, as in the method of Culler and Shurrager, but were supported in an upright position, in a standard conditioning stock. Because the animals were to be kept rather than disposed of on the day of the transection, there was no excision of the flexing muscles. The response to be conditioned was therefore a movement of the entire right-hind limb rather than the contraction of the exposed semitendinosus muscle.

In order to obtain records of the smallest possible movements of the limb, which were not in any way affected by movements of other parts of the animal's body, it was necessary to hold the spinal segment in an absolutely fixed position. Both sidesway and a fore-and-aft pulling of the hindquarters by the unparalyzed front feet had to be eliminated. A surgical device like the insertion of a drill-rod through the femurs was again impossible with chronic specimens. During all conditioning, therefore, the hindquarters of the animals were braced laterally between two rigid, yet adjustable, brackets which pressed firmly against the hips. The caudal end of the body was supported from above by leather straps encased in soft sponge-rubber tubes, which passed under each leg at its point of juncture with the body. Fore-and-aft movement was eliminated by an additional strap pulling the animal to the rear, which held the back of his head against a circular wooden collar or neck-piece. A pneumatic recording technique (9) was used to furnish kymographic records of the reactions of all four feet during the process of conditioning.

Surgery and maintenance.—To facilitate recovery from the operation, the injection of morphine sulphate or other depressing drugs was avoided, and ether alone was employed as the anesthetic agent. Under complete surgical anesthesia, the skin was incised for about three in. over the first and second, or the second and third, lumbar vertebrae. The muscle tissues were separated and the spinous and transverse processes exposed. These were then removed with rongeurs. The vertebrae were drilled with a dental burr, and the opening enlarged with rongeurs so as to expose the spinal cord for a distance of about 1½ in. In order to prevent the escape of spinal fluid, the cord was then ligated above and below the point where the transection was to be made. After the sectioning, which was done with microscopic scissors, the ends of the cord were observed to draw apart. Sulphanilamide was placed in the wound and the muscles and fascia were joined with interrupted silk sutures. Following the closing of the skin incision with tension sutures, one of the vertebrae of the animal's tail was drilled, and a metal pin was inserted for use in supporting the hindquarters during his maintenance.

Upon the completion of the operation, the *S* was transferred to a maintenance form where he was permitted to come out of the ether. His chin was placed upon a shelf, his chest rested in a broad canvas hammock, and his hindquarters were supported by means of the tail pin which allowed them to swing free and clear of all contacts. When not in the conditioning stock, the spinal preparations spent most of their time in this maintenance form (8). It was found to be a great aid in keeping them clean and in a healthy state during their post-operative life. The surgery was checked by autopsy upon the completion of the experimental work.

Training sequence.—The plan of the present investigation was to condition the animals *before the transection*, and to follow through with further training subsequent to the surgery. The

stimulus pattern was the same both before and after the operation. There could thus be no question about the fact that each individual in the normal state was fully capable of assimilating the training. It would also be possible by this method to measure retention effects, if any persisted after the operation. The first trials in spinal conditioning were given as soon as the animal was alert and receptive to food. The time elasping between the end of the operation and the beginning of the spinal trials usually ran from two to three hours. The reason for this hurry-up procedure was to duplicate as nearly as possible the technique of conditioning the *acute* spinal animal. In the present study, 100 conditioning trials administered in a single experimental session about two hours in length, were given in this way on the day of operation.

After the first experimental session, the *S* was returned to the maintenance form and allowed to rest. The second group of 100 spinal trails was administered 48 hours later. Training was continued in groups of 100 trials which were given in most cases on alternate days, until a total of 1000 trials of spinal conditioning had been amassed by each *S*. These trials were therefore distributed over a period of approximately three weeks following the transection. Retention effects—if there were any—from one experimental session to the next, should be readily observable by this method.

If an animal suffered from surgical shock, or if he failed to remain in a good condition sufficiently long to attain the quota of 1000 spinal trials, his results have not been included in the present record. There were four dogs which met the criteria outlined—three males and one female. All were active and healthy at the end of the training program. Each of these animals had previously received 400 original conditioning trials by the shock-shock method, in the control experiments already mentioned. Just before the transection, each was given enough additional pre-operative training to bring his conditioned behavior to a high level of efficiency (100 percent in three of the four cases) on the day before the operation.

The total number of original, pre-operative, and spinal conditioning trials received by each of the *S*s is given in Table I.

TABLE I

NUMBER OF TRIALS GIVEN TO EACH SUBJECT

Subject	Normal Conditioning			Spinal Conditioning	Total
	Original	Pre-operative	Both		
F 54	400	160	560	1000	1560
M74	400	300	700	1000	1700
M77	400	120	520	1000	1520
M79	400	100	500	1000	1500
All four	1600	680	2280	4000	6280

QUALITATIVE RESULTS

Spinal flexion responses.—Since the responses recorded in the present investigation were movements of the entire limb, while those observed by Culler and Shurrager were, in effect, the reactions of a nerve-muscle preparation, it may be argued that the present recording technique was, by comparison, crude and insensitive. Such an argument might be valid were it not for the fact that muscle-twitch reactions, like those reported by the former investigators, were present and recordable in the intact hind limbs of our own spinal animals. Since the response was not measured at the muscles, but was taken through the secondary medium of the bones of the leg and foot, there is no denying the fact that it was not as easy to observe,

and was in some cases doubtful or questionable. It must be remembered, however, that Shurrager and Culler found conditioned phenomena in less than 50 percent of the cases examined—or in 98 out of 219 preparations (22).

This muscle-twitch, or instantaneous jerk of the right-rear limb, occurred with varying frequency in all four of the spinal animals of the present study. It was small in amplitude—one to two mm.—and did not change according to any consistent trend as training proceeded. The latency of the response was 200 milliseconds or less and the duration was of similar magnitude. Since the left-rear leg flexed also as a direct response to the conditioned stimulus, the net result when the right leg twitched, was a bilateral flexing response, the amplitude of which was about four in. in the left foot and very small in the right. The flexing twitch of the right-hind foot might be elicited on any trial by either or both of the conditioned shocks which were sent to the left-rear foot. The nature of this twitch reaction in the right-rear member of the spinal animal may be seen in the lower record to the left in Fig. 1.

The crossed-extension reflex.—Undoubtedly because of the locus of the conditioned stimulus employed in the present study, there was also present a second and conflicting type of response to that stimulus, in three of the four spinal *S*s. This was an *extension* of the right-rear member as a result of the conditioned shocks to the left foot—in other words, the crossed-extension reflex. The extension response was, of course, opposite and antagonistic to the flexing twitch and rarely appeared on the same trials with it. There were likely to be several trials in which the extension response occurred, followed by a series in which the muscle-twitch predominated. The duration of the extension reaction was entirely different from that of the twitch, since the former would frequently continue for several full seconds after the stimulus had ceased. Extension reflexes of this sort in the right-hind foot of the spinal animal are shown in the lower record to the right in Fig. 1.

Are these responses spinal CRs?—That these two opposing movements are not exclusively spinal phenomena, but can occur with even greater facility in the normal organism, is shown by the two top records in Fig. 1. All of the records reproduced in this figure are records of the same experimental animal, M79. A comparison of the two records to the left in Fig. 1 shows the difference between the pre-operative and post-operative muscle-twitches. A comparison of the two records to the right shows the difference between the pre-operative and post-operative extension responses.

In the first of the two trials shown in the upper left-hand record, the pre-operative flexions, it will be noted, are of large amplitude.

On the third trial, however, the twitches are of the same order of magnitude as those obtained in the spinal state. It may also be seen from the upper left-hand record that the muscle-twitch reaction can occur in *either* of the rear limbs when the contralateral member is shocked. Not only are there twitch responses in the right leg when the left foot receives the conditioned stimulus, but twitch responses

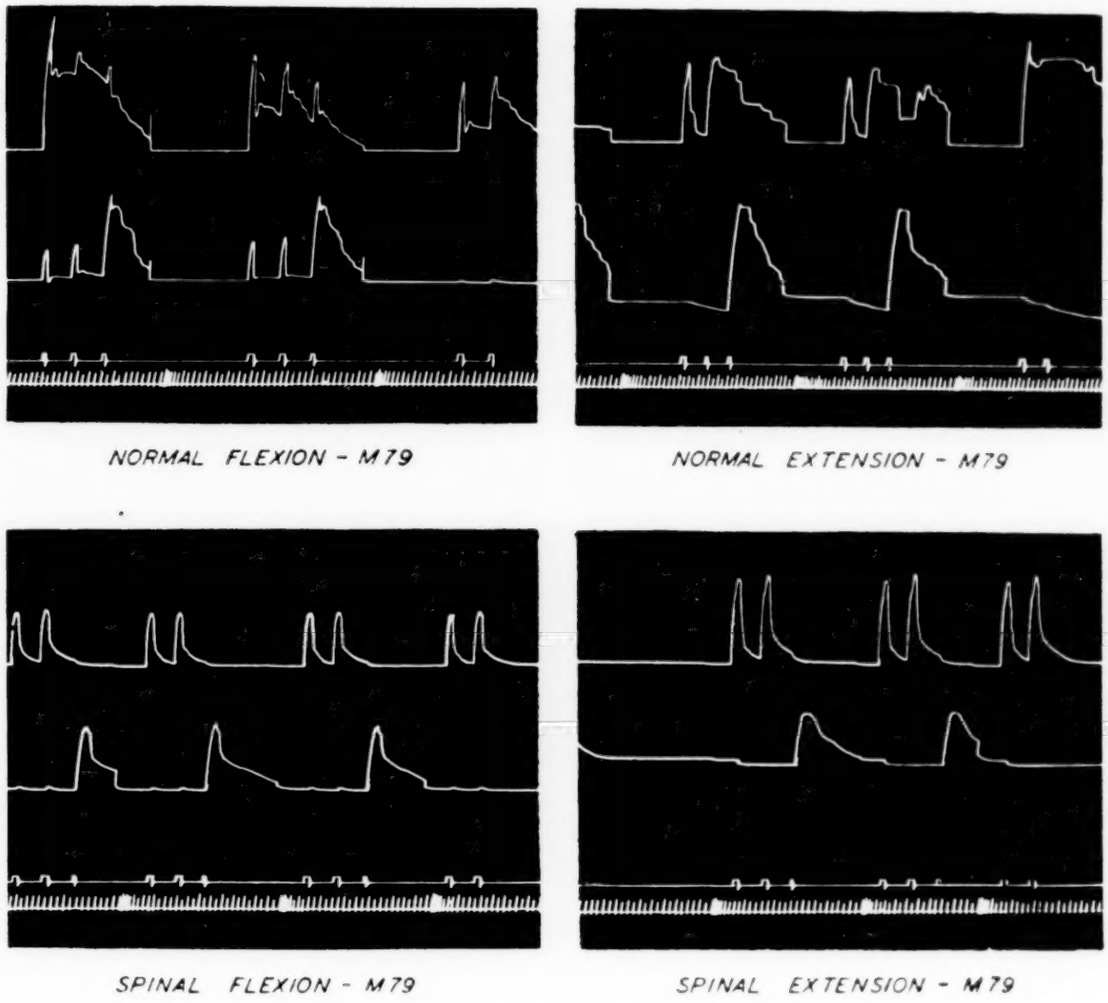

FIG. 1. Reproductions of kymographic records of muscle-twitch (flexion) and crossed-extension responses in the same animal, M79, before and after transection of the spinal cord. The top line in all records is the activity line of the left-rear foot. The line second from the top shows the activity of the right-rear foot. The first two breaks on any trial in the third, or stimulus, line indicate the sending of conditioned a.c. shocks to the left-rear foot, and the third break on any trial *in the same line* is the unconditioned d.c. shock to the right-rear foot. Time is in units of .2 sec. Three trials are shown in each of the records excepting the lower left-hand record in which there are four. In the final trial on each record the unconditioned stimulus to the right-rear foot has been omitted.

The upper left-hand record shows flexions or twitches in both rear feet in the intact animal, from stimulation of the contralateral member. In the third trial of this record the muscle-twitch responses in the right-rear leg are about identical in size to those which occur after transection, as shown in the lower left-hand record.

The two right-hand records show depressions or crossed-extension reflexes in the right-rear foot, from stimulation of the left-rear. The post-operative or spinal extensions (lower right-hand record) are much smaller than those of the intact organism (upper right-hand record).

can also be seen in the left leg as a result of the unconditioned d.c. shock to the right. It is apparent, therefore, that a shock of sufficient intensity to elicit a full flexion reflex of four in. in amplitude in either hind limb, may also elicit a muscle-twitch in the contralateral limb.

Flexion and extension responses of this sort are likely to occur very early in the training of the normal dog—often before the appearance of a genuine CR. There are grounds for regarding the extension response at least as an incomplete and inadequate adjustment to the stimulating situation. That they bear no resemblance whatever to the usual anticipatory conditioned flexion response, can be seen from the left-hand record in Fig. 2, which shows full-fledged pre-operative CRs, again from subject M79.

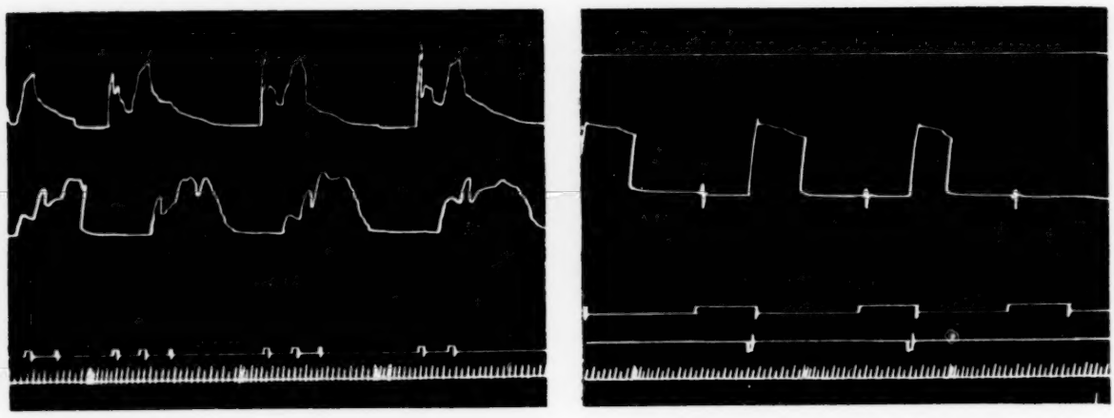

FIG. 2. The left-hand record shows normal anticipatory CRs of M79 before transection, when the stimulus pattern is the same as that used after transection. Reflex responses of M79 have previously been given in Fig. 1. The right-hand record of Fig. 2 shows the muscle twitch response in the right-rear foot of a normal dog, F40, when the conditioned stimulus is an a.c. shock two full sec. in length which is delivered to the homolateral *front* foot.

The two top lines in each of these records trace the activity of the left- and right-rear limbs respectively, and the bottom line gives time in units of .2 sec. In the left-hand record the conditioned and unconditioned stimuli are both recorded on the same line as in the records of Fig. 1. In the right-hand record the conditioned stimulus (delivered in this instance to the right-*front* foot) shows on the line third from the bottom, and the unconditioned stimulus to the right-*rear* foot shows on the line second from the bottom. The last trial to the right of each record is a trial in which the unconditioned stimulus has been omitted.

Both the extension reaction and the muscle-twitch or instantaneous flexing movement have been observed in normal dogs in our laboratory for some years past. A report by Pronko and Kellogg (**16**) in 1942, deals specifically with the latter phenomenon as an aspect of conditioned behavior in dogs. The muscle-twitch was regarded at that time as a function of the shock-shock conditioning procedure, and has only very rarely been noted in buzz-shock conditioning. It was obtained in the right-rear foot of dogs in response to a conditioned electric shock sent to either one of the front feet, of

sufficient intensity to elicit a full flexion in the foot which received it. In some instances, the muscle-twitch was "observable not only in the member to be conditioned, but in the other legs as well" (**16**, p. 233). The nature of this phenomenon when the shock is delivered to the homolateral front foot, and its comparison with the *bona fide* CRs of M79 before transection, can be seen in Fig. 2.

By combining the findings of Pronko and Kellogg with the results obtained from the control groups of the present investigation, together with the observations of Culler and Shurrager, we now feel that it is reasonable to conclude that the muscle-twitch response can be produced in the dog by stimulating any of the extremities—and probably in fact by stimulating any part of the body whatever. Our own observations have shown that it will occur in one or more of the limbs by electrical stimulation of some other limb, and Culler and Shurrager have shown that it will occur in the leg (or in the flexing muscles) upon electrical stimulation of the tail. Quite likely it is a part of the general response of the organism to any abrupt stimulation. The presence of something similar in the respiration curve—in the form of a sudden gasp when a shock is received—has long been recognized both in animals and humans.

The muscle-twitch reaction of the leg itself was originally regarded by Pronko and Kellogg as a new and distinct sort of conditioned response and was called by them a Type-T CR, to differentiate it from the normal anticipatory lift which they called a Type-L CR (**16**). In the light of the evidence which has appeared since, we now believe it incorrect to speak of the twitch reaction as genuinely conditioned at all. Its proper place in the scheme of behavior, according to the interpretation of the writers, will be more fully discussed at a later point in the present paper.

Quantitative Results

Frequency of responses.—Even though it is unlikely that the flexion or extension reactions of our spinal animals are truly conditioned, the data have nevertheless been analyzed according to traditional conditioning procedures. The purpose of such an analysis is, of course, to show in just what respects the responses are, or are not, like those of the conditioned organism. A frequency tabulation of both flexion and extension reactions has therefore been made, for the training series as a whole, as well as for various divisions of the training series.

The gross frequencies for both types of spinal reflexes in 1000 conditioning trails per animal are given in Table II. That these frequencies are very much lower than those reported by Culler and Shurrager, may be due to any one of the following factors. (a) The

total number of conditioning trials administered in the present study —although not greatly different from the total number of trials given by Culler and Shurrager—was divided among four animals instead of among a much larger group. (b) The data for all *Ss*—both 'successful' and 'unsuccessful'—are presented here. (c) The recording techniques were different.

TABLE II

RELATIVE FREQUENCY OF RESPONSES BEFORE AND AFTER TRANSECTION

Subject	Percent Pre-operative CRs		Percent Spinal Responses in 1000 Trials		
	Last 100 Trials	Last 20 Trials	Flexion Rs	Extension Rs	Flexion plus Extension
F 54	69	100	3.1	0.0	3.1
M74	79	80	8.4	34.2	42.6
M77	69	100	.2	3.9	4.1
M79	93	100	7.3	8.1	15.4
All four	78	95	4.8	11.6	16.3

For comparative purposes the frequencies of the CRs given by each animal in the last 100 trials—as well as in the last 20 trials—of his pre-operative conditioning, are also shown in Table II. Even when the spinal flexion and spinal extension responses are thrown together—a somewhat dubious procedure, but one which is done in this instance in order to present the results in the most favorable light—the frequencies after the operation show no meaningful relationship to those before. There is no indication from these data that preoperative learning has any influence whatever upon the spinal reflexes following transection. But perhaps one should not expect a carry-over from the pre-operative to the post-operative state, since the CRs before transection are so completely different from the reflexes which occur afterwards (compare Figs. 1 and 2).

A fractionating of the frequency data into groups of 100 trials is carried out in Tables III, IV, and V. If actual learning has taken place within a single experimental session, there should be some evidence of its retention or transfer to the next following session. The retention of several varieties of conditioned responses in normal dogs has been adequately demonstrated. When conflicting or antagonistic behavior is involved, as in the present case, the response retained is always that which was elicited in the last, or most recent, training trials (11). In the total of 40 100-trial experimental periods which were given to the four spinal animals of the present study, there were 36 intervals of approximately two days each, over which this phenomenon might be expected to show itself. Yet it was always con-

TABLE III

PERCENT FREQUENCY OF SPINAL FLEXION RESPONSES IN EACH 100 TRIALS

Subject	Groups of 100 Trials										All Trials (1000)
	1	2	3	4	5	6	7	8	9	10	
F54	0	1	2	0	0	0	18	4	1	5	3.1
M74	0	10	34	8	3	16	2	3	4	4	8.4
M77	1	0	0	0	1	0	0	0	0	0	.2
M79	9	2	26	2	0	30	4	0	0	0	7.3
All four	2.5	3.3	15.5	2.5	1.0	11.5	6.0	1.8	1.3	2.3	4.8

TABLE IV

PERCENT FREQUENCY OF SPINAL EXTENSION RESPONSES IN EACH 100 TRIALS

Subject	Groups of 100 Trials										All Trials (1000)
	1	2	3	4	5	6	7	8	9	10	
F54	0	0	0	0	0	0	0	0	0	0	0.0
M74	0	41	39	38	20	3	41	59	43	58	34.2
M77	0	0	5	6	25	3	0	0	0	0	3.9
M79	0	0	1	64	0	0	1	0	15	0	8.1
All four	0.0	10.3	11.3	27.0	11.3	1.5	10.5	14.8	14.5	14.5	11.6

TABLE V

PERCENT FREQUENCY OF FLEXION PLUS EXTENSION RESPONSES IN EACH 100 TRIALS

Subject	Groups of 100 Trials										All Trials (1000)
	1	2	3	4	5	6	7	8	9	10	
F54	0	1	2	0	0	0	18	4	1	5	3.1
M74	0	51	73	46	23	19	43	62	47	62	42.6
M77	1	0	5	6	26	3	0	0	0	0	4.1
M79	9	2	27	66	0	30	5	0	15	0	15.4
All four	2.5	13.5	26.8	29.5	12.3	13.0	16.5	16.5	15.8	16.8	16.3

spicuously absent. The frequencies in Tables III, IV, and V offer little dependable evidence of an increase—or, for that matter, of a decrease—between adjacent experimental sessions. M74, who was the most active animal of the group, is the only one whose results show any sort of continuity.

Response curves.—A further fractionating of the data into groups of 20 trials is given in Fig. 3. In the upper portion of this figure the flexion and extension responses have been plotted separately. The points on the graphs give the relative frequency of responsiveness,

within successive groups of 20 conditioning trials, for the four animals taken together. Each point, therefore, represents the flexion or extension frequencies within a total of 80 trials.

The bottom curve of Fig. 3 combines the data which are separately charted in the upper part of the figure. In this instance each point on the graph includes the sum of the flexion and extension responses for the four *S*s, within each group of 20 trials. It will be noted that the frequency rises irregularly to a general level of 15 to 20 percent

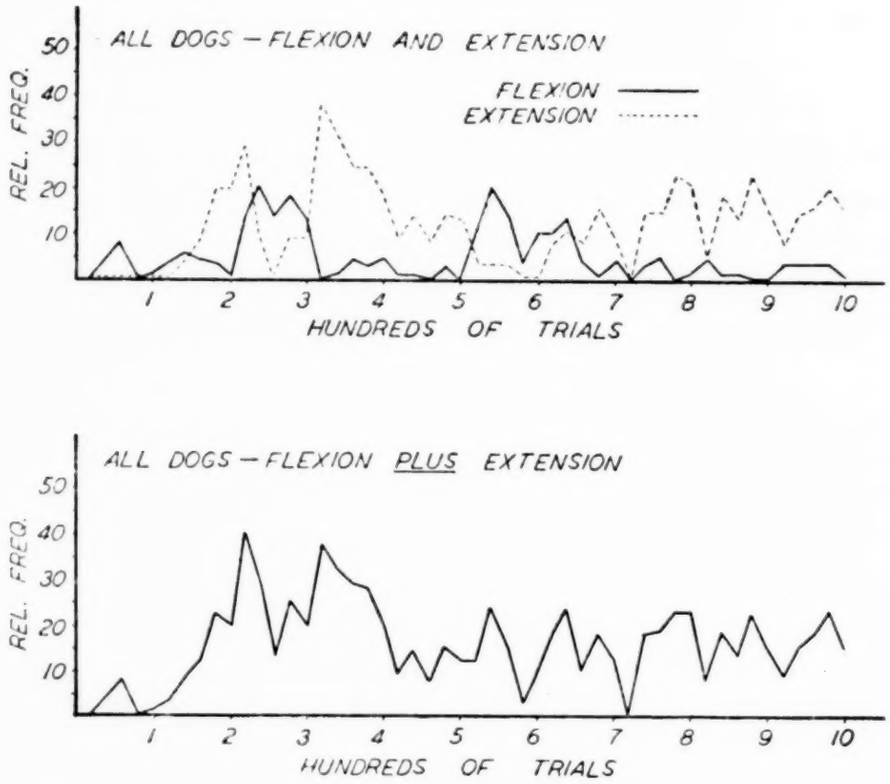

FIG. 3. Relative frequency graphs of the responsiveness of the four spinal animals taken as a group, for 1000 conditioning trials. The data are plotted in units of 20 trials. Each point on the graphs therefore represents a total of 80 trials for the four *S*s. In the upper part of the figure the flexion and extension responses have been separately plotted. In the lower part of the figure the two responses are totalled, without regard for their qualitative difference.

Aside from the fact that the graphs all start at zero, they bear little resemblance to the typical learning or conditioning curve. The extremely low frequencies and great irregularity are especially significant.

where it remains—with many fluctuations—for the entire 1000 trials. Aside from its initial rise, this graph shows but little resemblance to the typical conditioning curve. The lack of continuity or consistency from one group of trials to the next, as well as the very low level of efficiency, make it hard to evaluate as a possible curve of learning.

Since the data in Tables II, III, IV, and V show M74 to be the most promising animal in terms of frequency of responsiveness, separate response curves for this subject are presented in Fig. 4.

Even though M74 was unusually active, his behavior in no way suggests true learning or conditioning. Indeed the record is striking in its large recessions and intermittent high frequencies, which are masked or smoothed in the average curves of Fig. 3. A general trend for the flexion response to decrease with repetition, independently of the muscle-twitch reaction, may be seen in this figure.

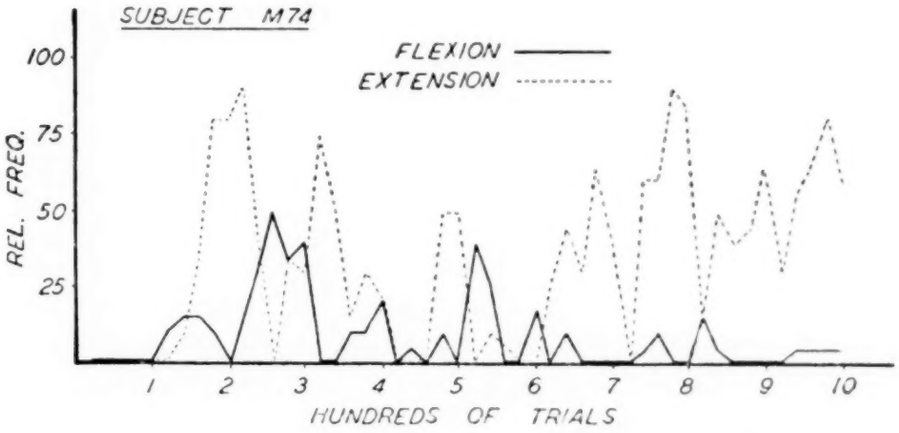

FIG. 4. Frequency curves for M74, the most active spinal animal of the group. In this case the intermittent high frequencies and numerous recessions make the behavior look even less like learning than in the smoothed or averaged curves of Fig. 3. Note that there seems to be a tendency for the flexion responses to drop out with repetition. As in the case of Fig. 3, each point on the graphs is the relative frequency of responsiveness for 20 trials.

DISCUSSION AND INTERPRETATION

It seems clear that this compound behavior—involving, as it does, two antagonistic reactions to the same stimulus—cannot be adequately explained as conditioning. There is no way that we can see to fit it properly into the conditioned-reflex formula.

As far as the extension response is concerned, this is undoubtedly a function of the point of stimulation of the conditioned stimulus. Had the conditioned stimulus been delivered to the tail or to some portion of the spinal segment other than the contralateral limb, the crossed extension-reflex would probably not have occurred.

With regard to the muscle-twitch, however, the situation is different. The twitch reaction we believe to be independent of the point of stimulation upon the body surface. We are also inclined to regard the muscle-twitch as a striped-muscle component of the startle pattern (**12**). If this is the case, it would occur in response to any abrupt or sudden stimulus. As described by Landis and Hunt, the startle pattern in man involves "blinking of the eyes . . . contraction of the abdomen, and *bending of the knees*" (**12**, p. 21, italics ours). This means that the complete reaction includes as a major element a rapid jerk or contraction of the flexing muscles of the legs,

which is quite likely the same as the muscle-twitch observed in the spinal dog.

"The response is very rapid and follows sudden, intense stimulation. It is a basic reaction, not amenable to voluntary control, is universal, and is found in Negroes as well as whites, infants as well as adults, in the primates and in certain lower animal forms" (**12**, p. 21). That elements of it can apparently be elicited in the spinal as well as in the normal animal only points to the overall significance of the pattern as a whole in the behavior of the organism.

The original response to S1.—It is our opinion further that the muscle-twitch response is probably a part of the natural reaction *to the conditioned stimulus alone.* In other words, the twitch response to the conditioned stimulus is a portion of the original response which that stimulus—if of sufficient intensity—will elicit by itself, without unconditioned reinforcement of any sort. The conditioned stimulus, however, must be sufficiently strong to produce this reaction; if weak, there may be no observable response. In the present experiments, it will be remembered, the conditioned stimulus was regulated so as to produce a full flexion in the limb to which it was delivered. The reaction which resulted, included not only the basic flexion of the stimulated member, but in many cases a simultaneous muscle-twitch in other portions of the body.

The scheme in Fig. 5, adapted from a diagram originally published by Hilgard (**6**), may serve to clarify the present interpretation. In the complete conditioning situation, three qualitatively distinct responses are often observable. There is the original unconditioned reflex, R1, to the unconditioned stimulus, S1. There is the natural or original response, R2, to the conditioned stimulus, S2. And there is the CR which usually appears before S1, in anticipation of S1. The CR in many cases is quite different both from R1 and from R2, and is not to be confused with either one of them.

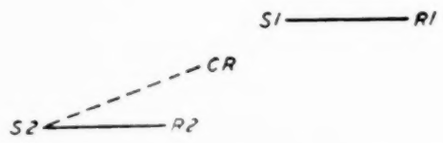

FIG. 5. Conditioned response diagram slightly modified from Hilgard (**6**). S1, in the present instance, is the unconditioned d.c. shock to the right-rear foot, and R1 is the direct flexion reflex which it produces. S1 may also produce a muscle-twitch in the non-shocked members (as in the upper left-hand record of Fig. 1). S2, in the present instance, is the series of two a.c. shocks per trial, which are sent to the left-rear foot. R2 (the response to S2 on any given trial) is two full flexions of the left-rear limb plus twitches in the contralateral member of the spinal preparation, and probably additional twitch-contractions in other limbs (or muscle groups) of the intact animal. The CR, which is to be differentiated both from R1 and from R2, is present as a response which is anticipatory to S1 in the normal dog (as in the left-hand record Fig. 2), but is absent in the spinal organism.

These three responses have been separately recorded by Hilgard (**6, 7**) in the conditioning of the eyelid or wink, response. The puff of air (S1) produces an unconditioned wink of large magnitude. The light (S2) elicits a much smaller 'natural' wink. The CR, which develops later, is entirely independent of the other two responses and occurs midway between them in point of time.

In many experiments on conditioning, the existence of R2 is difficult to observe because of the stimulating and recording arrangements. When the conditioned stimulus is a sound of some sort, as in buzz-shock or possibly in salivary conditioning, its response, R2, may be a turning of the head or a pricking up of the ears in the experimental animal. In such instances R2 is likely to be overlooked or ignored by the experimenter.

In the Hilgard procedure, and in the present instance where electric shocks were used both as conditioned and unconditioned stimuli, R2, the original response to the conditioned stimulus, can be seen and recorded. It is the view of writers that the muscle-twitch reaction of the present experiments is an important component of this unlearned response, R2. The response to the conditioned stimulus, while primarily a four-in. flexion of the left-hind limb, would also spread to the right-hind limb where it was recorded in the form of the muscle-twitch. Instead of being conditioned or learned in any sense of the word, it may be regarded therefore as a sort of startle contraction, elicited by the electric shock of the conditioned stimulus itself.

Changes in the frequency of the response.—If such an interpretation is correct, it follows that changes in the frequency of occurrence of the muscle-twitch—and also changes in frequency of the crossed-extension reaction—must be considered as changes in the frequency of the reflexes themselves, rather than as conditioned phenomena. There already exists evidence for the modification of spinal reflexes with repetition, as far as extinction is concerned. Prosser and Hunter (**18**), working with spinal rats, have found that reflexes in response to stimulation of the leg and tail can be definitely reduced or extinguished altogether with repeated trials. Disinhibition takes place under the proper circumstances but the basic reflexes are unquestionably subject to change. Shurrager and Culler (**23**), also, in their studies of spinal phenomena have reported that the muscle-twitch in dogs can be completely and permanently extinguished.

On the positive side one may point to the observations of Sgonina (**19, 20**) which have already been cited, and also to those of Gos (**5**). After extinguishing the flexion reflex to a tactual stimulus in normal frogs, Gos found that it could be re-established by reinforcing with an electric shock. Although this looks at first glance like a good case of

conditioned behavior, there are two extenuating factors. (a) The 'conditioned' response was originally present in the first place without any reinforcement. It had to be extinguished before it was reinforced. (b) The same modification of behavior appeared more readily in some *spinal* frogs and in frogs with frontal lobectomy, than it did in normal frogs.

Whether phenomena like these should be classified as learning or conditioning is a point upon which there may be some room for discussion. Our own view is that better terms to apply to them would be something like *summation of stimuli*, or *reflex sensitization*, or the *Sensibilisierung* of Sgonina (**19**, **20**). There can be no doubt that behavior of this sort is difficult to elicit in the mutilated preparation or in an isolated organ of the body. It should be noted, as a result, that there seems usually to be a considerable percentage of the cases in which it cannot be demonstrated at all.

Summary and Conclusions

If spinal conditioning is possible, it means that rudimentary learning can occur in an isolated portion of the organism after direct neural connections to the cerebrum, the basal ganglia, the cerebellum, and the medulla oblongata have been severed. The object of the present research was to explore the possibility of such conditioning in the hind limbs of the dog, following complete transection of the spinal cord. The investigation was patterned, in the main, after the work of Culler and Shurrager, although it differed from their experiments in important characteristics, any one of which may have influenced the results.

A. Essential features of the *procedure* were as follows:

1. The animals in the present experiment were *chronic* (rather than *acute*) preparations which were maintained for approximately three weeks following the transection. In this way it was possible to permit recovery from the operation and to increase considerably the number of trials received by any individual. The difficulty of keeping chronic spinal animals, on the other hand, greatly reduced the number of cases which could be observed. In the present study there were, as a result, four preparations which met the criteria of health and number of training trials which were set—although both operated and non-operated controls were used.

2. All four dogs were completely conditioned *before transection* to the point where they gave a high percentage of conditioned flexion responses. After the operation, 1000 additional conditioning trials were administered to each individual. The post-operative trials

were given in 10 separate sessions of 100 trials each, spaced on alternate days, following the transection.

3. The conditioned stimulus consisted of two (instead of three) a.c. electrical impulses, separated by a one-sec. interval. These shocks were delivered to the left-rear foot (rather than to the tail). One sec. following the last of the conditioned shocks on any given trial, the unconditioned stimulus—a short d.c. shock—was sent to the right-rear foot, which therefore served as the member to be conditioned. The intensity of both the conditioned and the unconditioned shocks was regulated so as to elicit a maximum flexion reflex of four in. in the two limbs which received the stimuli.

4. During their training, the animals of the present study were rigidly held by mechanical means in an upright (not an inverted) position.

5. And, finally, in our procedure the movement of the entire limb was recorded (instead of the movement of the isolated semitendinosus muscle, which was observed by the former investigators).

B. The principal *results* which came to light in the present experimental situation are summarized below:

1. There were two distinct response tendencies evident in the limb to be conditioned.

2. The first of these was a slight muscle-twitch or flexing jerk of the leg. The amplitude of the muscle-twitch response was one to two mm. and its latency was 200 milliseconds or less. This reaction is believed to be basically the same phenomenon as the conditioned muscle-twitch observed by Culler and Shurrager in the exposed semitendinosus muscle.

3. The second of the responses observed in the present animals was an extension of the member to be conditioned—in other words, the crossed-extension reflex. The occurrence of this second response is no doubt a direct result of the fact that the conditioned stimulus in the present instance was sent to the contralateral foot.

4. These two reactions were mutually inhibitory and were rarely observed on the same trial. One was likely to occur for a number of trials after which it would be superseded by the other. Each was fluctuating and irregular, even though the stimulating conditions remained the same.

5. Both the flexing twitch and the extension reaction have also been observed in our laboratory in normal dogs. The jerk or twitch response will occur in one or more of the legs when electrical stimuli are delivered to any of the other limbs. Unlike the crossed-extension reflex, the muscle-twitch would seem to be independent, therefore, of the point of delivery of the shock stimulus.

6. Neither of these movements bears any resemblance to the *bona fide* flexion CR given by an intact animal. If the conditioned and unconditioned stimuli are electric shocks, both may occur (in the normal dog) well before a genuine CR has been established.

7. The frequency of occurrence of the twitch and the extension reactions of the four spinal animals—when taken separately, or when added together—gave graphs which were unrecognizable as learning or conditioning curves.

8. There was no evidence of retention either (a) from pre-operative to post-operative trials, or (b) over the 36 temporal intevals of about two days which separated the 40 spinal training sessions of the four preparations.

9. Aside from these facts, it should be clear on *a priori* grounds that the compound behavior which occurred—involving as it did two antagonistic responses to the same stimulus—cannot be satisfactorily fitted into the conditioned-reflex formula.

C. The major *interpretations* of these findings are as follows:

1. It is probable that the muscle-twitch reaction to the conditioned stimulus can be elicited without unconditioned reinforcement of any kind, providing the conditioned stimulus itself is strong enough. That a synchronous pairing of the conditioned and unconditioned stimuli is not necessary to call it out, is clearly demonstrated by its occurrence in the present experiment where these two stimuli were never synchronously paired. In addition to the withdrawal of the left-hind foot which was the primary reaction to the conditioned stimulus in the present instance, the response to a sudden electric shock will frequently spread to other body parts. If recorded in the contralateral member of the spinal preparation, it will appear as the muscle-twitch reaction. The twitch response is interpreted, therefore, as a part of the natural or basic response *to the conditioned stimulus alone.*

2. This twitching response is probably also a striped-muscle component of the startle pattern which can be produced by any abrupt tactual or electrical stimulus, regardless of the point of delivery of that stimulus upon the surface of the body. Its presence as a gasp in the respiration curve has long been recognized.

3. Changes in the frequency of either the twitch or the extension reaction, which occur with repetition, are therefore considered as simple variations in reflex activity.

4. It is suggested that terms like *summation of stimuli* or *reflex sensitization*, would be more suitable to apply to changes of this sort, than to regard them as instances of genuine conditioned behavior.

5. The changes in the frequency of muscle-twitch responses in normal dogs which were formerly thought by Pronko and Kellogg to be conditioned (**16**), are now interpreted as cases of reflex sensitization. It is possible—in the light of the present findings—that the phenomenon of spinal conditioning reported by other investigators should also be classified in the same way.

(Manuscript received March 23, 1946)

REFERENCES

1. CULLER, E. Observations on the spinal dog. *Psychol. Bull.*, 1937, **3**, 95–110.
2. CULLER, E., & METTLER, F. A. Conditioned behavior in a decorticate dog. *J. comp. Psychol.*, 1934, **18**, 291–303.
3. CULLER, E., & METTLER, F. A. Observations upon the conduct of a thalamic dog. Hearing and vision in decorticated animals. *Proc. Soc. exp. Biol.*, N. Y., 1934, **31**, 607–609.
4. GIRDEN, E., METTLER, F. A., FINCH, G., & CULLER, E. Conditioned responses in a decorticate dog to acoustic, thermal, and tactile stimulation. *J. Comp. Psychol.*, 1936, **21**, 367–385.
5. GOS, M. Le psychisme de la moelle épinière. *Bull. Soc. Sci. Liége*, 1932, **4**, 95–97.
6. HILGARD, E. R. The nature of the conditioned response: I. The case for and against stimulus-substitution. *Psychol. Rev.*, 1936, **43**, 366–385.
7. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York, Appleton-Century, 1940.
8. KELLOGG, W. N. A method for the maintenance of chronic spinal animals. *J. exp. Psychol.*, 1946, **36**, 366–370.
9. KELLOGG, W. N., DAVIS, R. C., & SCOTT, V. B. Refinements in technique for the conditioning of motor reflexes in dogs. *J. exp. Psychol.*, 1939, **24**, 318–331.
10. KELLOGG, W. N., PRONKO, N. H., & DEESE, J. Spinal conditioning in dogs. *Science*, 1946, **103**, 49–50.
11. KELLOGG, W. N., & WOLF, I. S. The nature of the response retained after several varieties of conditioning in the same subjects. *J. exp. Psychol.*, 1939, **24**, 366–383.
12. LANDIS, C., & HUNT, W. A. *The startle pattern.* New York, Farrar and Rinehart, 1939.
13. LASHLEY, K. S. *Brain mechanisms and intelligence.* Chicago, Univ. Press, 1929.
14. LEBEDINSKAIA, S. I., & ROSENTHAL, J. S. Reactions of a dog after removal of the cerebral hemispheres. *Brain*, 1935, **58**, 412–419.
15. METTLER, F. A., METTLER, C. C., & CULLER, E. Effects of total removal of the cerebral cortex. *Arch. Neur. Psychiat., Chicago*, 1935, **34**, 1238–1249.
16. PRONKO, N. H., & KELLOGG, W. N. The phenomenon of the muscle-twitch in flexion conditioning. *J. exp. Psychol.*, 1942, **31**, 232–238.
17. PRONKO, N. H., & KELLOGG, W. N. Reflex mechanisms and the ease of conditioning. *Amer. J. Psychol.*, 1942, **55**, 371–384.
18. PROSSER, C. L., & HUNTER, W. S. The extinction of startle responses and spinal reflexes in the white rat. *Amer. J. Physiol.*, 1936, **117**, 609–618.
19. SGONINA, K. Vergleichende Untersuchungen über den bedingten Reflex. *Zool. Anzeiger*, 1939, **125**, 248–251.
20. SGONINA, K. Vergleichende Untersuchungen über die Sensibilisierung und den bedingten Reflex. *Z. Tierpsychol.*, 1939, **3**, 224–247.
21. SHURRAGER, P. S., & CULLER, E. Phenomena allied to conditioning in the spinal dog. *Amer. J. Physiol.*, 1938, **123**, 186–187.
22. SHURRAGER, P. S., & CULLER, E. Conditioning in the spinal dog. *J. exp. Psychol.*, 1940, **26**, 133–159.
23. SHURRAGER, P. S., & CULLER, E. Conditioned extinction of a reflex in the spinal dog. *J. exp. Psychol.*, 1941, **28**, 287–303.

REACTIVELY HETEROGENEOUS COMPOUND TRIAL-AND-ERROR LEARNING WITH DISTRIBUTED TRIALS AND TERMINAL REINFORCEMENT [1]

BY CLARK L. HULL

Institute of Human Relations, Yale University

Introduction

In 1939 Chester J. Hill published an experimental study (3) concerning heterogeneous [2] compound trial-and-error learning with terminal reinforcement.[3] His apparatus was a four choice-point linear maze with four choices at each choice point. His trials were slightly massed in that his animals were given five trials in immediate succession each day, for a total of 10 days. This study showed among other things a tendency for both ends of the maze, i.e., the first and fourth choice points, to be easier for the animals to learn than either of the two intermediate choice points when the criterion of difficulty taken was first-choice errors. The exact position of the maximum varied from the second to the third choice point according to the criterion used, as shown in Fig. 1. Hill attributed the special difficulty in the learning at the second and third choice points of the maze to the convergence towards the middle of the series of the competitional interference by the other (heterogeneous) reactions being learned at the same time.

In 1943 Thompson and Dove published a related experimental

[1] Throughout this experiment Ruth Hays acted as laboratory technician. She also scored the original graphic records and tabulated the primary data. The writer is indebted to Harry G. Yamaguchi for the calculation of the statistical reliabilities. The apparatus represented in Fig. 2 was constructed by our mechanician, K. G. Ogren.

[2] Reactively heterogeneous compound trial-and-error learning is distinguished from reactively homogeneous compound trial-and-error learning in that by the former a *different* reaction is required at each choice point whereas by the latter the *same* reaction is required at each choice point. In the heterogeneous variety, there must be the same number of choices at each choice point as there are choice points themselves (see Fig. 2), whereas in the homogeneous variety the number of choices may be fewer. In addition to the two pure forms of compound trial-and-error learning just contrasted there are numerous mixed forms in which homogeneous and heterogeneous phases are combined. Among the latter may be mentioned single and double alternation learning investigated by Hunter (7, 8), Warden (12, 13), and others.

[3] *Terminal* reinforcement is to be contrasted with *serial* reinforcement in that in the former primary reinforcement (e.g., the giving of food) occurs only following the correct performance of the *last* act of the chain, whereas in the latter, reinforcement occurs at once after the performance of *each* act of the chain. The present article is the second of a series involving various combinations of the above variants in compound trial-and-error learning and chain reactions (1) now completed or nearing completion in the laboratories of the Institute of Human Relations (6).

investigation in compound trial-and-error learning (**10**), the results of which led them to question Hill's interpretation of his experiment. In particular they suggest that the characteristic outcome of Hill's study, i.e., the error gradients sloping generally from the middle downward towards the two ends—the so-called bi-directional gradient—was an artifact produced by the degree of massing of trials involved in his experimental procedure. They assume that the effect of Hill's feeding his animals at the end of his first, second,

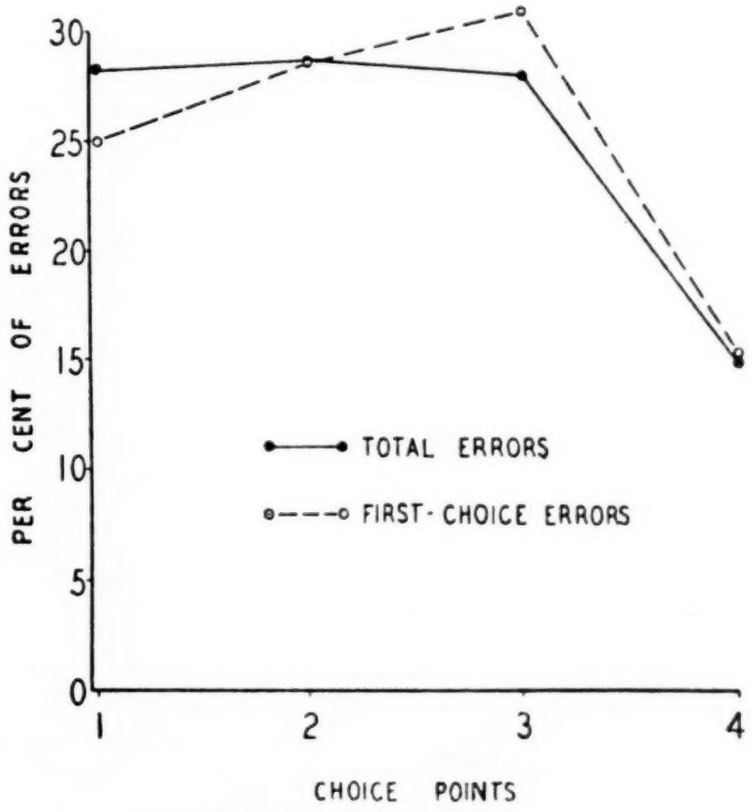

FIG. 1. Graphic representation of the percent of errors made at the several choice points by Hill's animals when the basis of computation is (1) total errors and (2) first-choice errors. The basis of the percent computation in each case is the total number of errors of the specified type performed at all four choice points combined. Note that the maximum percent of total errors occurs at choice point 2, whereas that of first-choice errors occurs at choice point 3. (Plotted from Hill's published data, 3.)

third, and fourth trials would perseverate in accordance with Thorndike's hypothesis (**11**, p. 55) and produce a special facilitation of the learning at the early part of the second, third, fourth, and fifth trials which immediately followed the feeding on each day. In this connection they remark (**10**, p. 455):

> One variable which has not been controlled in these [Hill's] studies is that of massed and distributed practice. Is it possible to demonstrate that the uni-directional gradient becomes a bi-

> directional one under massed practice? Hill's study certainly indicates as much. If this is the case, can we still explain both gradients by Hull's goal gradient hypothesis . . .?

The present study was carried out primarily to test this suggested possible effect, i.e., the production of a uni-directional gradient by purely distributed practice under the conditions otherwise operative in the Hill experiment.

Apparatus, Procedure, and the Scoring of Results

In general the *Ss* used and the apparatus employed in the present investigation were as exactly like those used by Hill as was feasible under the circumstances. Hill's apparatus is shown as Fig. 2. The only apparatus change made by us was to place a 25-watt, frosted electric bulb at a distance of 10.5 in. above each of the five narrow passageways marked E in the drawing of the maze. The bulbs were enclosed in wooden boxes 3.5 in. in each dimension, which restricted the illumination mostly to the maze. The shade covering the window of the experimental room was always drawn. The only significant deviation from the technique employed by Hill was the substitution of strictly distributed trials—one trial per 24 hours except over week-ends when 48 hours separated the trails—instead of the mild form of combined massed and distributed trials employed by Hill.

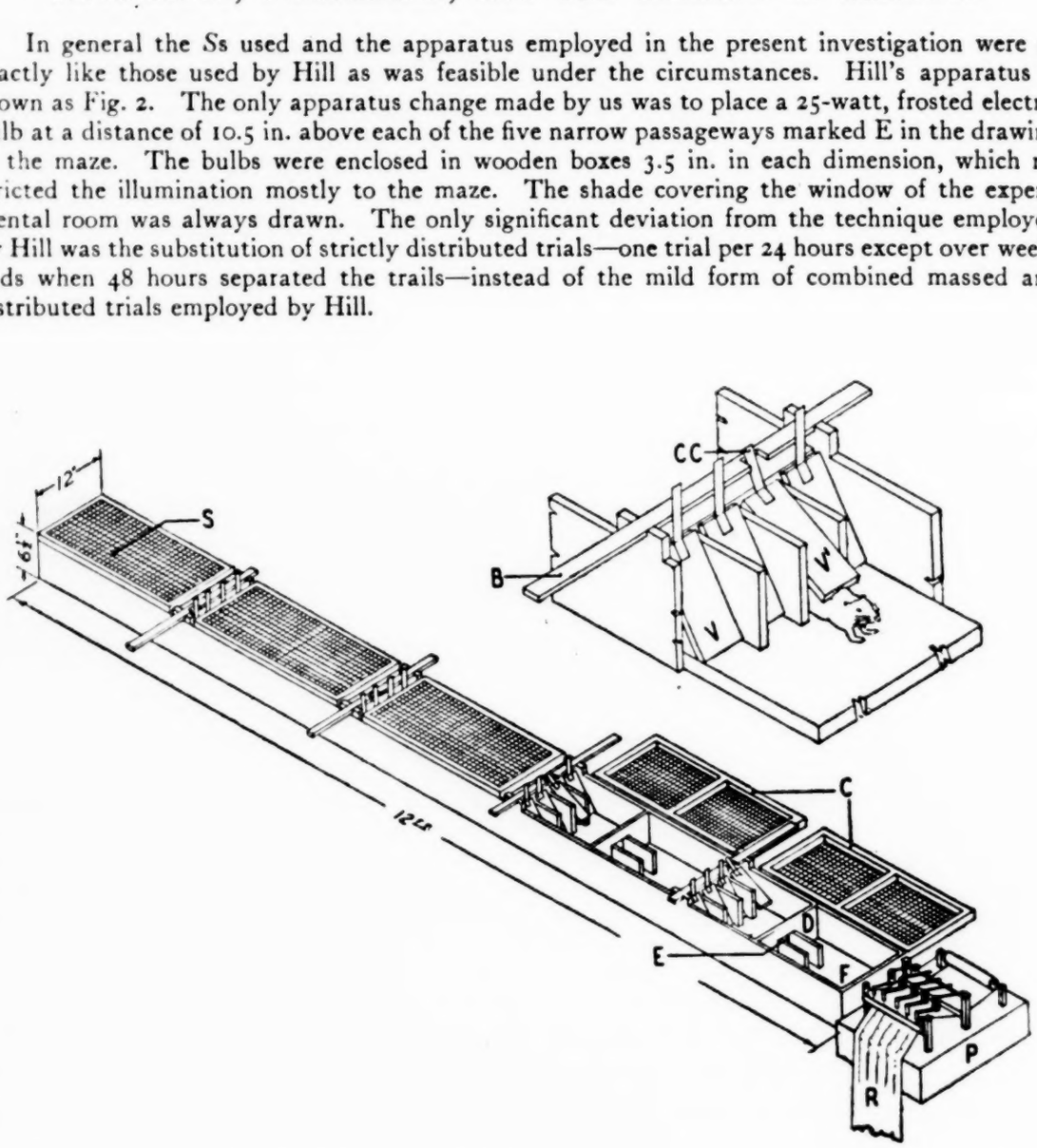

Fig. 2. Diagram of apparatus (lower figure) and one choice point (upper figure). Lower diagram: S, starting box; C, wire mesh frames covering each section of the maze; D, partition midway in each section; E, aperture to force rat to middle of maze; F, food compartment; P, constant-speed polygraph; R, waxed paper record. Upper diagram: B, bar to block incorrect doors (V); CC, contact in slot which allows correct door (V′) to be opened by pressure in goalward direction. (Taken from Hill, 3.)

A change was also made in the scoring of the records. In the frequent cases where a rat pushed several times in succession at a door which was locked, before trying another door, thus making several distinct successive notches of the same error on the polygraph record, Hill counted only a single error unless a period of 10 sec. or an attempt on a different door intervened between such acts. In the present experiment all such distinct notches were counted as errors without exception, in order to facilitate comparison with Thompson and Dove's results. This change in the method of scoring the records naturally increases considerably the 'total' error scores in the present report over those of Hill's study, especially in the early trials.

The animals in the present experiment were male albino rats purchased from the Albino Farms, Redbank, New Jersey. Their diet was made up of Gaines' dog biscuit supplemented once a week by lettuce and codliver oil. They were run under a 23-hour food deprivation. In the training trials the animal was put into the maze at S (Fig. 2). At each choice point during the main training trials three of the valves were locked and one was free to lift at the animal's pressure (see insert, V'), a different valve at each choice point, exactly as in Hill's study. When the animal reached point F he found four pellets of food made from a wet mash consisting of pulverized dog-chow biscuit, wheat flour, and water. The pellets were about $\frac{1}{4}$ in. in diameter and lay on a slip of white paper 3 by 5 in. in size. After an animal had eaten the pellets it was picked up by hand and returned to its living cage. All 24 sequential combinations of the valves of the maze were employed, three animals on each, which involved 72 animals in all, exactly as in Hill's study. One set of 24 combinations was completed before the next was begun. All animals alike were given 50 trials, after which the training was discontinued. Accordingly, each animal necessarily made a total of 50 correct choices (each often preceded by errors) at each of the choice points of the maze. The sequence of the running of each group of 24 animals was constantly varied to prevent any learning of the maze pattern from odor cues left by the tracks of the preceding animal; this part of the technique differed from that employed by Hill and may have made the learning slightly more difficult.

Results

Two methods will be followed in reporting the choices of the animals. By the first method *all* of the errors are scored, and by the second, only the first choice at each choice point is tabulated. The first method is of special interest because Thompson and Dove report their experimental results in this manner; the second is of importance because it offers unique opportunities for analysis owing to the presumption that the first choice made is the strongest reaction potentiality existent in that animal at that particular instant.

The total errors of all kinds made by our 72 animals on their 3600 trials ($72 \times 50 = 3600$) are shown analytically in Table I. In order to facilitate comparison with Thompson and Dove's results as well as with those of Hill, the values in Table I are retabulated in Table II in such a way as to show the total number of errors made at each choice point. A study of Table II shows that during the first five trials there is some tendency for the number of errors to fall from the first choice point, as Hill's results did, but the tendency is not so clear as in Hill's experiment, since in Table II (column I) choice point 4 has a slightly greater number of errors (21.8 percent) than does choice point 3 (20.9 percent). The total errors, including percents, for all stages of training are given in the final column of Table II. These results are also represented graphically in Fig. 3 to permit comparison with Thompson and Dove's Fig. 3 (10, p. 453).

TABLE I

TOTAL ERRORS OF ALL 72 ANIMALS AT ALL POINTS OF THE MAZE AT FIVE STAGES OF TRAINING. The valves are numbered according to the choice point at which the valve in question is correct, i.e., not locked. The valves which are numbered the same as the choice point have no entries in this table, since no errors can be made at those choices. This table is comparable (except as noted in the text) with the non-italicized portion of Hill's Table III (3, p. 574).

Choice point	No. of valve	Set of 10 trials					Total of 50 trials
		I	II	III	IV	V	
1	1						
	2	1136	297	218	220	207	2078
	3	1075	303	185	162	155	1880
	4	1169	257	208	184	135	1953
2	1	1008	324	238	190	130	1890
	2						
	3	1190	348	320	278	218	2354
	4	1133	415	356	259	216	2379
3	1	744	324	210	143	101	1522
	2	865	374	262	199	181	1881
	3						
	4	847	645	584	427	338	2841
4	1	735	225	130	99	75	1264
	2	871	201	76	63	79	1290
	3	949	288	164	126	105	1632
	4						

TABLE II

TOTAL ERRORS MADE BY ALL ANIMALS AT THE SEVERAL CHOICE POINTS AT FIVE STAGES OF TRAINING, IRRESPECTIVE OF THE TYPE OF ERROR MADE. The parentheses following the several error numbers show the percent of errors made at that stage of training at the choice point in question. This table is derived from Table I and is therefore approximately comparable with Hill's Table II (3, p. 573), but is differently arranged.

Choice point	Sets of 10 trials					Total of 50 trials
	I	II	III	IV	V	
1	3380 (28.8)	857 (21.4)	611 (20.7)	566 (24.1)	497 (25.6)	5911 (25.7)
2	3331 (28.4)	1087 (27.1)	914 (31.0)	727 (30.9)	564 (29.1)	6623 (28.8)
3	2456 (20.9)	1343 (33.6)	1056 (35.8)	769 (32.7)	620 (32.0)	6244 (27.2)
4	2555 (21.8)	714 (17.8)	370 (12.5)	288 (12.2)	259 (13.3)	4186 (18.2)

A comparison of the present results with those of Hill shows no evidence that the strictly distributed practice here employed changed the error distribution from that yielded by the slightly massed trials

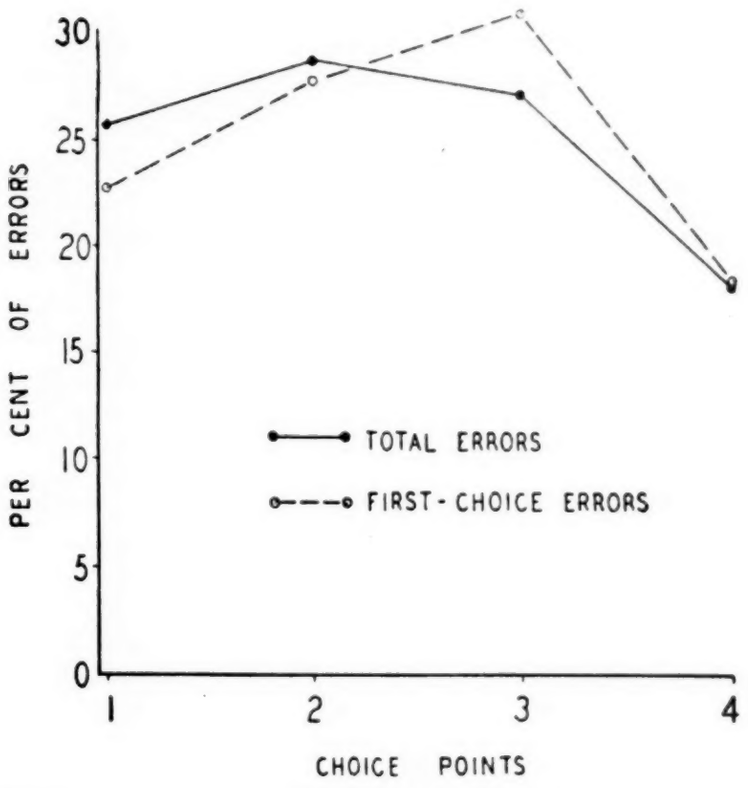

FIG. 3. Graphic representation of the percent of errors of the indicated type made at the several choice points by the animals of the present study when the basis of computation is (1) total errors and (2) first-choice errors. The basis of percent computation in each case is the total number of errors of the specified type performed at all four choice points combined. Note that the maxima of the two types of errors fall at the same choice points as in the Hill study. These graphs have been plotted from data shown in Tables II and IV respectively. In comparing Figs. 1 and 3 it must be recalled that 'total errors' were scored somewhat differently in the two studies.

of Hill's experiment. Neither does the present study give any confirmation of the increase in a tendency to a single start-to-goal gradient, as anticipated by Thompson and Dove. Actually, Hill's error results show a greater tendency to a falling gradient from start to goal than do the present ones. Indeed, Hill's corresponding percentages do not differ so very much from those of Thompson and Dove in a formal sense, since a rise of less than one percent in total errors at Hill's choice point 1 would have made his gradient of total errors technically a uni-directional gradient. They were: 28.2%, 28.7%, 28.0%, and 14.9%.[4] However, to be exact, the Hill study and the present one agree in showing a maximum of total errors at the second choice point and a minimum at the fourth choice point. Hill does not report any statistical reliabilities. The critical ratios based on the standard errors [5] of the differences between the mean

[4] It must be recalled in this connection that Hill's data at this point are not strictly comparable with our own because they were scored in a slightly different manner.

[5] The standard errors of the several differences were calculated by means of the formula,

$$\sigma_{\text{diff.}} = \sqrt{\sigma_{M_1}^2 + \sigma_{M_2}^2 - 2r\,\sigma_{M_1}\,\sigma_{M_2}}.$$

The six correlation coefficients average .466.

number of errors made at each point and those made at each of the other points, each followed by the statistical reliability (in parenthesis), are:

I vs. II = 1.512 (.935)		
I vs. III = .578 (.718)	II vs. III = .676 (.750)	
I vs. IV = 3.164 (.999)	II vs. IV = 4.581 (1.000−)	III vs. IV = 4.639 (1.000−)

When we pass from the consideration of the total errors to the total first-choice errors we encounter a rather different relationship, exactly as Hill did. The detailed picture of the first-choice behavior at five stages of training is yielded by Table III. Perhaps the most

TABLE III

ANALYTICAL TABLE SHOWING THE FIRST CHOICES MADE BY THE ANIMALS AT EACH CHOICE POINT AT FIVE STAGES OF TRAINING. The numbers representing the correct choices are set in bold-faced type. The valves are numbered according to the choice points at which the valve in question is correct, i.e., not locked. This table is to be compared with Hill's Table IV (3, 577).

Choice point	No. of valve	Sets of 10 trials					Total of 50 trials	Percent choices at each valve
		I	II	III	IV	V		
1	1	**256**	**405**	**458**	**482**	**485**	**2086**	**57.9**
	2	173	121	100	101	114	609	16.9
	3	144	91	84	63	63	445	12.4
	4	147	103	78	74	58	460	12.8
2	1	146	94	55	56	48	399	11.1
	2	**234**	**326**	**360**	**394**	**433**	**1747**	**48.5**
	3	188	144	150	153	129	764	21.2
	4	152	156	155	117	110	690	19.1
3	1	124	72	56	45	31	328	9.1
	2	172	88	46	41	53	400	11.1
	3	**255**	**275**	**281**	**335**	**388**	**1534**	**42.6**
	4	169	285	337	299	248	1338	37.1
4	1	134	87	44	51	31	347	9.6
	2	140	68	34	32	45	319	8.9
	3	181	108	100	88	86	563	15.6
	4	**265**	**457**	**542**	**549**	**558**	**2371**	**65.9**

significant single view obtainable of this rather complex set of data, and one to be contrasted with that of total errors, is the total number of first-choice errors as shown in detail in the next-to-last column. When these errors at each choice point are combined and the percent each sum is of the whole is calculated, we have Table IV. The percent of total first-choice errors alone made at the several choice points is represented by the broken-line curve in Fig. 3. A comparison with Hill's analogous results as shown in Fig. 1 indicates a remarkably close agreement between the two studies at all points, with the maximum in both falling at choice point 3.

TABLE IV

THE TOTAL FIRST-CHOICE ERRORS AT ALL FIVE STAGES OF LEARNING MADE AT THE SEVERAL CHOICE POINTS ARE HERE SUMMARIZED FROM TABLE III. The percent values are represented graphically in Fig. 3.

Choice point	Number of errors	Percent errors
1	1514	22.7
2	1853	27.8
3	2066	31.0
4	1229	18.4

The critical ratios and the statistical reliabilities (in parenthesis) of the differences between the mean number of first-choice errors at the various choice points upon which are based percent successes shown in Table IV and Fig. 3 (broken line) are as follows: [6]

I vs. II = 3.069 (.999)		
I vs. III = 4.115 (1.000−)	II vs. III = 1.367 (.915)	
I vs. IV = 2.562 (.995)	II vs. IV = 5.282 (1.000−)	III vs. IV = 5.685 (1.000−)

Here it may be seen that the probability that the obtained differences between the minimal point (choice point 3) and each of the others were due to chance alone was less than one percent in two cases, and less than nine percent in one. The fact that the results of Hill's study parallel those of the present one so closely, appreciably supports and strengthens the one weak reliability, i.e., that between choice points 2 and 3.

In order to give the reader as full and prompt a view as possible of the relationship of the several aspects of first-choice behavior, the percent values shown in the last column of Table III are represented graphically in Fig. 4. For purposes of comparison the analogous relationship, of the success scores *only*, existent in Hill's data is represented as a strictly comparable curve which appears in the upper part of the figure. There a glance at the two success curves reveals that the parallelism of the two studies is very close indeed, despite the fact that the present study shows an appreciable and consistently smaller percent of correct first-choice responses. As already pointed out, the slower learning of our animals may very well have been caused by the systematic variation of the fresh tracks of recently run animals which would otherwise have been available as the stimulus cues upon which to base choices. Even so, it may be observed that such additional margin of disagreement as exists consists in a somewhat greater accentuation in the present experiment of the weakness of choice point 3 as related to choice points 1, 2, and 4 when

[6] The mean of the correlation coefficients is .366.

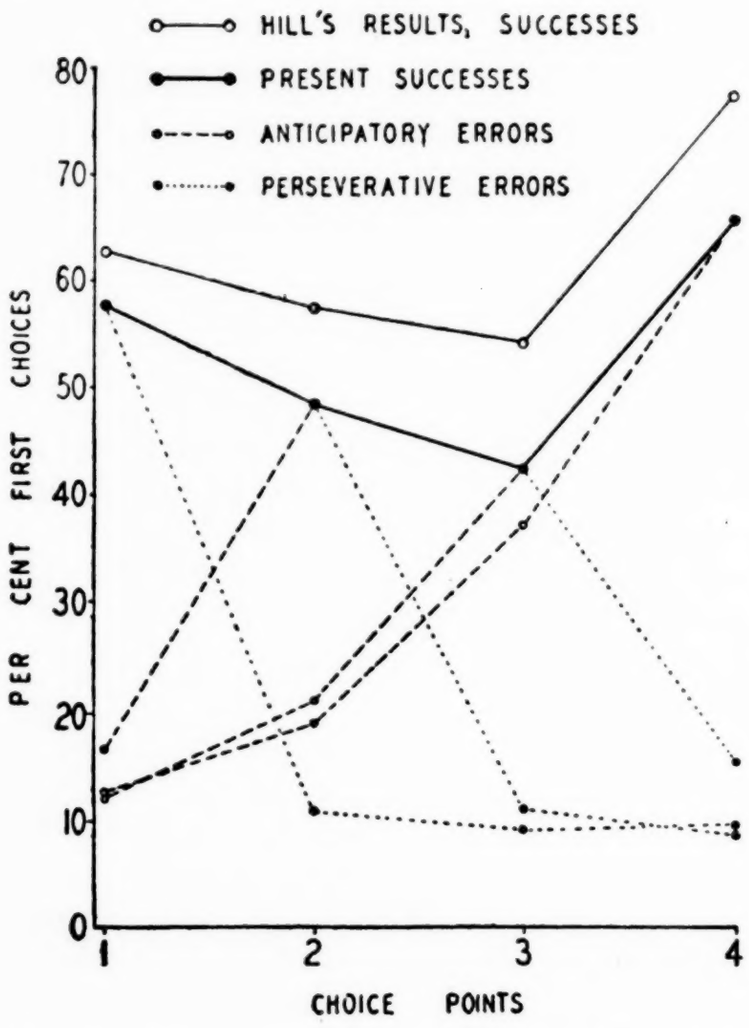

FIG. 4. The lower graph represents the percent distribution of the correct and the erroneous first-choice reactions of the present experiment as *made at a given choice point.* Note the progressive fall in the relative frequency of the occurrence of reactions correct at any given choice point, as the distance from that choice point increases. This portion of the graph is plotted from the data given in the last column of Table III. For purposes of comparison, a graph of the analogous percent successful reactions only in Hill's experiment is given in the upper part of the figure.

compared with similar points of Hill's data.[7] This means that the strictly distributed practice of the present experiment is associated with an even greater exaggeration of the so-called bi-directional gradient tendency than was the slightly massed practice of Hill's experiment. Thus, from this additional angle the Thompson-Dove hypothesis quoted above is not substantiated.

We now leave the comparison of the present study with that of Hill, together with that of Thompson and Dove, as a matter of

[7] It is to be noted that the first-choice graphs of Hill's and the present study here under comparison (Fig. 4) appear to be the inverse of the graphs in Figs. 1 and 3. But even this inverse comparability is only approximately true, because Figs. 1 and 3 are calculated on the basis of the sum of the errors alone *at all four choice points*, whereas the percents involved in Fig. 4 are based on the total number of reactions, both errors and correct choices, performed *at each single choice point.*

primary concern, and pass to the analysis of the internal relationships of the various types of errors to the correct choices. The basic facts for this comparison are presented in easily grasped form in Fig. 4, but in considerably greater detail in the body of Table III.

An examination of the lower portion of Fig. 4 in connection with Table III shows in general a progressive diminution in the probability of the tendency of the animals to choose a valve at one choice point in the same position as one which is correct at another choice point. Consider the series of valves which are in the same relative position as the correct valve at the fourth choice point. Taking the total first choices as a basis for comparison, we find that the percents of choices of this valve at the respective choice points are (final column, Table III):

I	II	III	IV
12.8	19.1	37.1	**65.9**

Similarly in the case of the valve which is correct at the third choice point, we have the following percents of choices at choice points earlier in the series:

I	II	III
12.4	21.2	**42.6**

Thus, the probability of making this erroneous choice grows less as the distance from the correct (and reinforced) choice point increases. In this connection it is to be observed that the errors in question are *anticipatory* or *antedating* reactions in the sense that the choice would have been correct if it had not occurred too early in the progress through the maze. Exactly parallel results are yielded by Hill's published data.

Next we observe that not only antedating reactions, but *perseverative* reactions occur, i.e., reactions which occur too late in the maze sequence to be correct. For example, taking the two reaction tendencies originating at the first and second choice points, which are analogous to the two just considered at the opposite end of the maze, we have,

I	II	III	IV
57.9	11.1	9.1	9.6
	48.5	11.1	8.8

With the exception of the slight increase at choice point 4, we find here, as with the antedating tendencies, a reduction in the probability of occurrence in the reaction as the distance from the point of reinforcement increases. Exactly parallel results are yielded by Hill's published data.

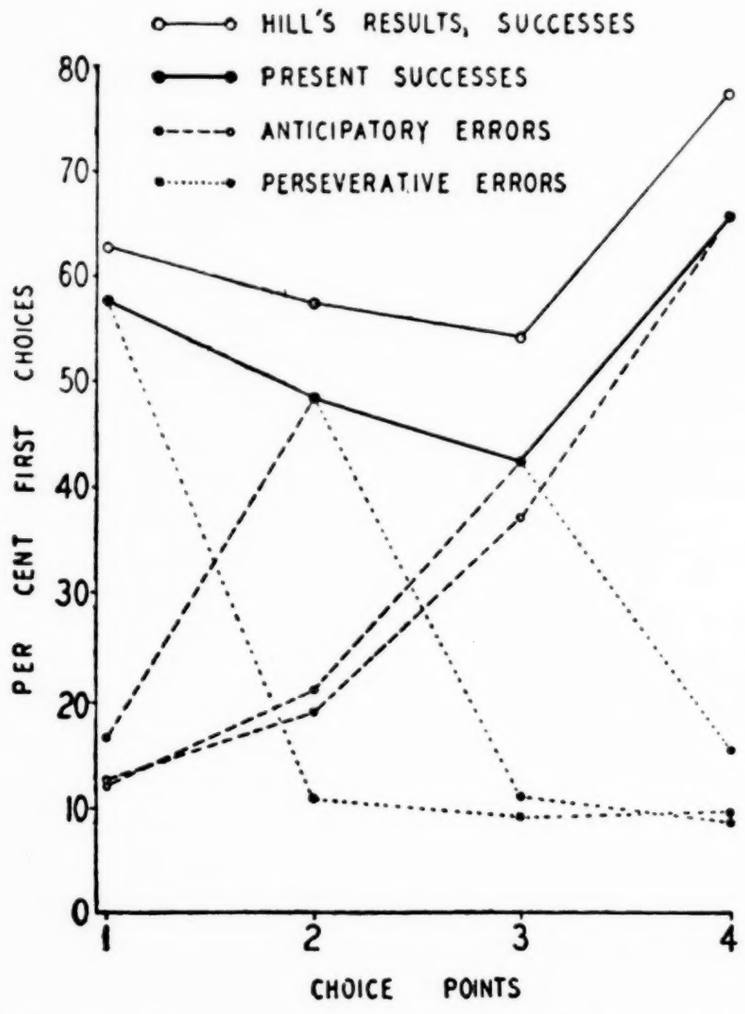

FIG. 4. The lower graph represents the percent distribution of the correct and the erroneous first-choice reactions of the present experiment as *made at a given choice point.* Note the progressive fall in the relative frequency of the occurrence of reactions correct at any given choice point, as the distance from that choice point increases. This portion of the graph is plotted from the data given in the last column of Table III. For purposes of comparison, a graph of the analogous percent successful reactions only in Hill's experiment is given in the upper part of the figure.

compared with similar points of Hill's data.[7] This means that the strictly distributed practice of the present experiment is associated with an even greater exaggeration of the so-called bi-directional gradient tendency than was the slightly massed practice of Hill's experiment. Thus, from this additional angle the Thompson-Dove hypothesis quoted above is not substantiated.

We now leave the comparison of the present study with that of Hill, together with that of Thompson and Dove, as a matter of

[7] It is to be noted that the first-choice graphs of Hill's and the present study here under comparison (Fig. 4) appear to be the inverse of the graphs in Figs. 1 and 3. But even this inverse comparability is only approximately true, because Figs. 1 and 3 are calculated on the basis of the sum of the errors alone *at all four choice points*, whereas the percents involved in Fig. 4 are based on the total number of reactions, both errors and correct choices, performed *at each single choice point.*

primary concern, and pass to the analysis of the internal relationships of the various types of errors to the correct choices. The basic facts for this comparison are presented in easily grasped form in Fig. 4, but in considerably greater detail in the body of Table III.

An examination of the lower portion of Fig. 4 in connection with Table III shows in general a progressive diminution in the probability of the tendency of the animals to choose a valve at one choice point in the same position as one which is correct at another choice point. Consider the series of valves which are in the same relative position as the correct valve at the fourth choice point. Taking the total first choices as a basis for comparison, we find that the percents of choices of this valve at the respective choice points are (final column, Table III):

I	II	III	IV
12.8	19.1	37.1	**65.9**

Similarly in the case of the valve which is correct at the third choice point, we have the following percents of choices at choice points earlier in the series:

I	II	III
12.4	21.2	**42.6**

Thus, the probability of making this erroneous choice grows less as the distance from the correct (and reinforced) choice point increases. In this connection it is to be observed that the errors in question are *anticipatory* or *antedating* reactions in the sense that the choice would have been correct if it had not occurred too early in the progress through the maze. Exactly parallel results are yielded by Hill's published data.

Next we observe that not only antedating reactions, but *perseverative* reactions occur, i.e., reactions which occur too late in the maze sequence to be correct. For example, taking the two reaction tendencies originating at the first and second choice points, which are analogous to the two just considered at the opposite end of the maze, we have,

I	II	III	IV
57.9	11.1	9.1	9.6
	48.5	11.1	8.8

With the exception of the slight increase at choice point 4, we find here, as with the antedating tendencies, a reduction in the probability of occurrence in the reaction as the distance from the point of reinforcement increases. Exactly parallel results are yielded by Hill's published data.

A third empirical principle derivable from the data just presented is that the rate of fall of both types of gradients grows progressively less, i.e., the gradient becomes progressively less steep, as the distance from the point of reinforcement increases. This may easily be observed from a study of the four gradients just considered, particularly as represented in Fig. 4. Exactly parallel results are yielded by Hill's published data.

A fourth empirical principle is that the initial fall of the perseverative gradient is markedly steeper than is that of the antedating gradient. This may most easily be observed by a comparison with the steepness of the corresponding initial segments of the gradients shown in Fig. 4. It is also observable by a comparison with the differences between the points of reinforcement and the next succeeding point in the antedating and the perseverative tendencies respectively. These differences are:

Antedating:	**65.9**–37.1	= 28.8
	42.6–21.2	= 21.4
	Mean	= 25.1
Perseverative:	**57.9**–11.1	= 46.8
	48.5–11.1	= 37.4
	Mean	= 42.1

Thus, 42.1 > 25.1; i.e., the initial segments of the perseverative gradients average considerably steeper than do those of the antedating gradients. Exactly parallel results are yielded by Hill's published data.

The fifth empirical principle which we derive from Table III concerns the slopes of the initial segments of the generalization gradients just considered as they are influenced by the learning process. We accordingly note the number of choices of the valve correct (i.e., reinforced) at choice point 1, at both choice points 1 and 2 during the first 10 trials, as 256 and 146 respectively. Dividing each of these by the total number of first choices at each point during 10 trials, 720, we have 35.5 percent and 26.2 percent respectively. The differences between these two percent values is 15.3, which yields a rough indication of the steepness of the slope of the gradient on this valve from the correct choice point to the first incorrect choice point, i.e., one choice point removed from the point of reinforcement. In an exactly analogous manner the corresponding differences in percent choices were calculated from the other columns of Table III for the remaining four stages of the training on this discrimination. These several steepnesses of slope as so quantified are represented graphically in the upper curve of Fig. 5, where the respective percent

differences are placed in the middle of the corresponding practice ranges.

Exactly analogous measures of the slope steepness of the corresponding antedating gradients (originating at the posterior end of the maze) at the several stages of the training were calculated and plotted as the lower curve in Fig. 5. A glance at these two curves reveals, in addition to the greater initial steepness of the perserevative gradient already derived as our fourth empirical principle, the fact that both the perseverative and the antedating gradients increase progressively in steepness as training increases. Exactly parallel results are yielded by Hill's published data.

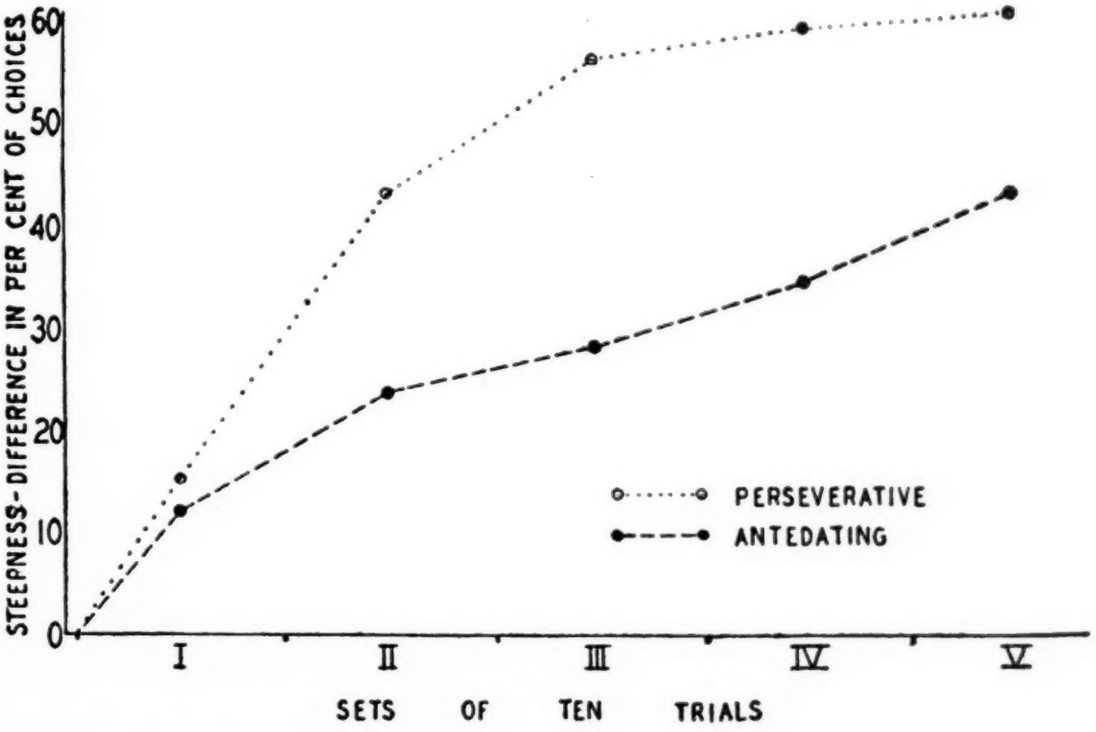

FIG. 5. The steepness in the generalization gradient as a function of the successive sets of 10 learning trials, placed in the middle of each set. Here steepness in gradient slope is taken as roughly indicated by the difference in percent of first choices due to one step of removal in the maze from the point of reinforcement.

DISCUSSION

Having presented the main empirical facts derivable from the experiment here under examination, we must pause to consider their scientific significance as to behavior dynamics in general. Thompson and Dove's challenge of Hill's interpretation of the latter's experimental results, and by implication of our own, serves to raise the vital question of the scientific principles involved and will be taken as a general point of departure, though our analysis necessarily goes much farther.

I. In this type of learning (see Fig. 2) it is evident at the outset that except for accidental imperfections in the apparatus unobservable to the investigator, the external stimuli at one choice point are exactly the same as those at every other. Nevertheless, in order to reach the food with a minimum of delay and work expenditure the organism must learn to execute a different movement sequence at the first choice made at each of the choice points. It follows that at the first choice at each choice point the percent choices of all four valves available must average approximately the same, i.e., chance or 25 percent. Thus the slope of the gradient at the first trial would be approximately $25\% - 25\% = 0$, as shown at the beginning of the curves of Fig. 5. However, as learning gets under way it follows that such a large and obvious sameness in the four stimulus components will necessarily produce a strong tendency of any special reaction potentiality learned at one choice point to generalize to every other choice point. This we have amply seen to be the case (Fig. 4).

Now, for any consistent success in error reduction to develop, the several choice points must somehow be discriminated. But discrimination must have a basis in stimulus differences at the choice points. Since the external stimulus components at the several choice points are presumably nearly identical, the *S*s must be dependent for differentiation almost entirely upon internal stimulus components. Moreover, these latter must consist of perseverative traces of stimuli occurring before the choice point is reached, notably including the proprioceptive stimuli arising from necessarily executed preceding movements. Thus, we comprehend a weak but genuine basis for the discriminal differentiation of the several choice points. It is evident from the preceding that the accumulation of characteristically distinctive stimulus traces as the animal progressively moves through the maze will make the stimulus composition at each choice point which is available for differentiation, progressively different from that of the others, the farther removed any choice point is from any other.

It follows from the above considerations that:

1. There must be a strong tendency to generalization of reaction tendencies from any point of reinforcement to all other choice points.

2. But after learning has appreciably occurred, the degree of the generalization tendency will not be complete, i.e., the generalization gradient will not be horizontal but will fall below the point of reinforcement.

3. As training progresses the differences in the stimulus compounds at the several choice points will be gradually conditioned positively to the reactions reinforced at these points and negatively to the reactions not reinforced at the same points.

4. This will cause:

(A) The progressive steepening of the gradients at the first segment as practice continues.

(B) The progressive flattening of gradients from one segment to the next, practice being constant.

II. It has been noted that the antedating generalization gradients are much less steep in fall than the perseverative generalization gradients. This means that the tendency to antedating generalization is much stronger and more conspicuous than is the tendency to perseverative generalization. This difference was noticed, in a sense, as far back as the time of Ebbinghaus. The most plausible reason for the difference in the two gradients seems to be found in the following considerations:

(A) Each stimulus trace spontaneously weakens in strength rather rapidly with the passage of time.

(B) Studies by Hovland (**4**, pp. 285, 288), Brown (**2**), and others show that the generalization from a weak stimulus to a strong one of the same sense field is strong, whereas that from a strong stimulus to a weak one is relatively weak.

(C) It follows from (A) and (B) that when a reaction is conditioned to a strong phase of a trace, and subsequently a later and consequently a weaker phase of the same trace impinges on the subject, the latter phase will have a relatively weak (perseverative) generalization tendency. This will cause the gradients under consideration to fall with a rapidity greater than would result from ordinary stimulus generalization (**5**, p. 185 ff.).

(D) It also follows from (A) and (B) that when a weak phase of a trace is conditioned to a reaction and subsequently an earlier, and therefore a stronger, phase of the same trace is encountered, a relatively strong antedating generalization tendency will occur. This will cause the antedating generalization gradient to fall slower than would result from ordinary stimulus generalization (**5**, p. 185 ff.).

(E) It follows from (C) and (D) that the initial segment of the perseverative generalization gradient must be steeper than the corresponding segment of the antedating gradient. Thus, the substance of the differing generalization gradients pointed out above seems to be explained to a first approximation, though there is doubtless more to the picture yet to be clarified.

III. It is evident that the two generalization gradients at the two ends of the series necessarily extend exclusively toward the rest of the series, i.e., toward the center of the maze. Moreover, the maximum section of this generalization falls at the choice point one step removed from the point of reinforcement and it is at a minimum at

the choice point farthest from the point of reinforcement. It follows from these considerations that the maximal accumulation of foreign generalization tendencies in a four-section maze will fall especially on the two central choice points. Also, (a) since the magnitude of the reaction potential to be generalized, because of the goal gradient principle (**5**, p. 135 ff.) is maximal, and (b) since the fall in antedating generalization gradients is much less rapid than is that of the perseverative generalization gradient, it follows that the maximal accumulation of this generalization effect will fall at the third rather than at the second choice point.

But since the reactions of a heterogeneous series are incompatible with each other, all generalization tendencies will naturally interfere with the evocation of the properly reinforced reaction tendencies to a degree which is a positive function of the magnitude of the total of such generalization. It follows that the maximum of generalization will cause a maximum of interference, i.e., of depression of the percent of correct responses, which normally would result from the feeding occurring at once after choice point 4. There will consequently result a serious distortion of the score of correct reactions, i.e, the goal gradient. Now, it is inevitable that *all* of the learning principles, e.g., the principle of heterogeneous generalization as well as that of the goal gradient, should operate in the Hill type of experiment, and no one exclusive of the others. It is therefore quite natural that even when the goal gradient is operative a sizable depression in the central regions of the maze should occur, and that the anterior end as well as the posterior end of the series should show a relatively higher mean score of successes than do the two central choice points. Naturally, because of the goal gradient, the terminal choice point shows a distinctly higher percent of successes than does the first choice point. Thus is reconciled the paradox of the distortion of the goal gradient mentioned by Thompson and Dove in the quotation cited above.[8]

IV. Incidentally the discussion just given is believed to clarify a persistent confusion in psychology inherited from the days when the so-called laws of 'primacy' and 'recency' were in vogue. As might naturally be expected wherever reactively heterogeneous elements are involved in any learning series, the generalized tendencies from the two ends, especially of rather short series, will inevitably depress the central portions of the series of whatever type of behavior, so

[8] No attempt at a systematic reconciliation of the present results with those reported by Thompson and Dove has been made. However, a conjecture is hazarded that their marked deviation from ours may be due in part to the relatively long maze employed as well as to the use of the continuous technique. A careful analysis of the generalization gradients involved could presumably throw light on an admittedly important problem. Time is lacking for the writer to do this at present.

that there will be a relatively slight advantage to the anterior end along with a very marked advantage to the posterior end. Actually such observations have frequently been made. The accompanying failure to recognize the operation of the behavioral principles involved as cited above, some time ago even led one writer favorably disposed to the recency-frequency concepts to regard the notion of the tendency to learn mazes "from the food-box backward (regressive order)" as 'exploded' (**12**, p. 252).

V. As a decisive means of testing the soundness of the reasoning given above, we offer the following simple deduction. Let it be supposed that an experiment be set up much as Hill's, but in such a way that all of the reactions in the maze instead of being heterogeneous are homogeneous, i.e., the *same* reaction is reinforced at every choice point. In this case generalization should occur exactly as assumed in the present experiment. However, in the homogeneous case the generalized tendencies would combine or summate with, rather than oppose, each other; this would have the effect that the central choice points would be stronger than would result from the action of the goal gradient alone. As a result it might easily happen that under these conditions the third choice point of a four choice-point series would show an even greater percent of successful choices than the fourth. Such an experiment would not be difficult of performance (**1**); indeed, an experiment already published strongly suggests the deduced outcome (**6**), though this was quite unnoticed by the experimenter who reported the investigation (**9**).

Summary

The Hill experiment, the interpretation of which was called in question by Thompson and Dove, has been repeated with 72 animals, strictly distributed trials being given. The results may be summarized as follows:

1. The present experiment gave the number of total errors made by the animals arranged from the maximum to the minimum in the following order of the maze choice points:

II, I, III, IV,

which agrees exactly with the order in this as reported by Hill.

2. The present experiment gave the number of the total first-choice errors made by the animals arranged from the maximum to the minimum in the following order of the maze choice points:

III, II, I, IV,

which agrees exactly with the order in this as reported by Hill.

3. It is accordingly concluded that the Thompson-Dove conjecture that the excess of errors in the middle of Hill's series rather than at the anterior end was due to the slight amount of massed practice involved, is without foundation.

A further analysis of the present results shows that:

4. The tendency for animals to perform errors antedating the points of reinforcement is a distinctive pattern of behavior.

5. This tendency grows progressively less as the error choice point is distant from the point of reinforcement.

6. The tendency for animals to perform errors perseverating from the point of reinforcement is also a distinct pattern of behavior and there is a similar weakening of the tendency as distance from the point of reinforcement is increased.

7. The slope of the first segment of the perseverative generalization gradient is noticeably steeper than is that of the corresponding part of the antedating gradient.

8. The slope of both gradients begins at the first trial at a practical horizontal and increases in steepness as training progresses.

9. A theoretical interpretation is proposed which purports to show that:

(A) The two types of stimulus generalization gradient generalize in such a way as to accumulate with special strength on the two central choice points.

(B) Because the generalized reactions are heterogeneous, this accumulation reduced the percent of correct reactions at the two central choice points below what would be yielded by the goal gradient operating alone.

(Manuscript received June 11, 1946)

References

1. Arnold, W. J. Simple reaction chains and their integration. Ph.D. thesis, 1946, on file in the Yale University Library.
2. Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.
3. Hill, C. J. Goal gradient, anticipation, and perseveration in compound trial-and-error learning. *J. exp. Psychol.*, 1939, **25**, 566–585.
4. Hovland, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.
5. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century Co., Inc., 1943.
6. Hull, C. L. Compound trial-and-error learning: pure forms. (Chapter VII of *Elementary theory of individual behavior*, now in manuscript.)
7. Hunter, W. S. The behavior of raccoons in a double alternation temporal maze. *J. genet. Psychol.*, 1928, **35**, 374–388.
8. Hunter, W. S. The sensory control of the maze habit in the white rat. *J. genet. Psychol.*, 1929, **36**, 505–537.

9. MONTPELLIER, G. DE. An experiment on the order of elimination of blind alleys in maze learning. *J. genet. Psychol.*, 1933, **43**, 123–139.
10. THOMPSON, M. E., & DOVE, C. C. A further study on the bi-directional goal gradient in the endless maze. *J. exp. Psychol.*, 1943, **33**, 447–456.
11. THORNDIKE, E. L. An experimental study of rewards. *Columbia Univ. Contr. Educ.*, 1933, No. 580.
12. WARDEN, C. J., & CUMMINGS, S. B. Primacy and recency factors in animal motor learning. *J. genet. Psychol.*, 1929, **36**, 240–256.
13. WARDEN, C. J., & HAMILTON, E. L. The effect of variations in length of maze pattern upon rate of fixation in the white rat. *J. genet. Psychol.*, 1929, **36**, 229–237.

REACTION TIME AS A MEASURE OF RETROACTIVE INHIBITION

BY LEO POSTMAN AND HAROLD L. KAPLAN

Harvard University

Retroactive inhibition has conventionally been measured in terms of decrements in retention due to an activity interpolated between learning and memory test. Tests of recall and recognition as well as the method of saving have been used for the demonstration of such decrements. Whatever the particular test procedure, the number of correct responses on the memory test is used to determine the degree of retroactive inhibition, usually by comparing the retention loss of an experimental group with that of a control group.

It is questionable whether measures in terms of retention loss bring out the full extent of retroactive inhibition. Retroactive inhibition refers to the interference of an interpolated activity with the traces of the original learning. It may also be regarded as negative transfer from the interpolated task to the initial learning task. Such interference or negative transfer need not always manifest itself in retention loss. Interference may not be serious enough to obliterate an item completely and yet detract from the efficiency of recall. Thus, overt intrusions from the interpolated activity often appear during the retention test but are quickly recognized as false by the *S* and replaced by correct responses (**5**). It may be assumed that in other cases such competition between correct and false responses does not issue in overt intrusions, but is solved by vicarious trial-and-error before the response is made. Provided such inhibitory effects are overcome within the period of time which *S* is allowed for a correct response on the memory test, they fail to affect the retention score. Retention scores, therefore, may not provide an exhaustive measure of all the detrimental effects of an interpolated activity on the memory for the original task.

To gauge the full extent of retroactive inhibition, a measure is needed which will supplement measurement in terms of retention scores. This method must be sensitive to such momentary inhibitions as do not manifest themselves in errors or failures to respond during the memory test. It was the purpose of the present investigation to demonstrate that reaction time for correct responses on the memory test can be used as such a measure. There are precedents

in the learning literature for the use of reaction time as a measure of associative strength (1, 2). Lengthened reaction time has also long been regarded as an index of inhibition, as, for example, in the free association technique (8, 340–367). Thus, reaction time measurement offered a promising avenue of approach to a precise determination of such retroactive effect as do not result in outright retention loss.

The Experiment

Subjects: Forty-four undergraduates at Harvard College served as *Ss*; 22 *Ss* served as an experimental group and 22 as a control group. All *Ss* were volunteers and were assigned to the two groups at random.

Experimental procedure: The paradigm conventionally used in experiments on retroactive inhibition was followed:

Experimental Group: Original Learning—Interpolated Learning—Relearning
Control Group: Original Learning—Relearning

The interpolated learning took place two min. after the end of the original learning series. In all cases relearning was 48 hours after original learning.

The original learning series consisted of twenty paried associates. Each pair consisted of unrelated nouns or adjectives (A—B). *E* read the total series once to *S*, with a one-sec. interval between the members of a pair and three sec. between successive pairs. After that, *E* presented the left-hand members of the pairs and *S* had to respond with the right-hand members within three sec. If he failed to do so, he was prompted. There was a five-sec. interval between successive pairs. There were five such learning trials during each of which the entire series was presented in a different prearranged random order.

The interpolated series also consisted of 20 paired associates (A—K). The left-hand member of each pair (A) was one used in the learning series but a different unrelated right-hand member (K) was substituted for the original right-hand member (B). This arrangement usually results in considerable retroactive inhibition (4, p. 490 f.). The learning procedure was in all respects the same as that followed during the original learning.

On the relearning trials *S* was presented with the left-hand members (A) and instructed to supply the original right-hand members (B). There were six relearning trials during each of which the entire series was again presented in random order.

On all trials *S*'s reaction time was measured by means of a voice-key arrangement. In pronouncing the stimulus word, *E* spoke into a microphone the activation of which completed an electric circuit and started a chronoscope. In giving the response word, *S* spoke into another microphone the activation of which broke the circuit and stopped the chronoscope. In this manner the reaction times of all *Ss* on all trials were recorded. If *S* failed to respond within three sec., *E* stopped the clock, prompted the *S*, and proceeded to the next stimulus word after a five-sec. interval.

Results

The results of the experiment fall into two parts: (1) measures of retention in terms or error scores and (2) measures of reaction time for correct responses. Retroactive inhibition can be measured in terms of (1) a comparison of the error scores of the experimental group and the control group, and (2) a comparison of the two groups with respect to their reaction times for correct responses.

Table I shows the average number of errors of the two groups on all trials. The differences between the two groups and the significance of these differences are also indicated. Under each experimental

TABLE I

AVERAGE NUMBER OF ERRORS FOR EXPERIMENTAL GROUP AND CONTROL GROUP ON ALL TRIALS. (Entries Include Differences Between the Groups and Tests of Significance of the Differences.)

	Average Number of Errors										
	Original Learning					Interpolation	Relearning				
Trial No.	Expt'l	Control	Diff.	*t*	*P*	Expt'l	Expt'l	Control	Diff.	*t*	*P*
1	15.95	15.95	0.0	0.0	1.00	13.14	12.18	9.04	3.14	3.42	<.01
2	12.18	14.00	1.82	2.00	.05	9.64	7.27	6.45	.82	1.01	.30
3	9.59	11.27	1.68	1.68	.10	6.04	5.45	4.77	.68	.70	.50
4	7.27	9.23	1.96	1.77	.10	5.23	4.14	3.18	.96	1.16	.30–.20
5	5.54	6.45	.91	.88	.40	3.73	2.95	2.91	.04	.04	>.90
6							1.91	1.82	.09	.02	>.90

condition, the error scores of both groups steadily decrease as learning progresses. During the original learning trials, the performance of the experimental group tends to be slightly superior, but on the last learning trial the two groups have converged sufficiently to make the difference between their mean error scores statistically insignificant. On the relearning trials, the performance of the control group is consistently superior to that of the experimental group. (The difference between the two groups is statistically significant only on the first relearning trial, however, when *raw* scores are considered.)

As a result of the interpolated activity, the performance of the experimental group ceases to be slightly superior and becomes consistently inferior to that of the control group. The extent of retroactive inhibition may be quantified by comparing the performance of the two groups on the last trial of the original learning with their performance on successive trials of relearning. The extent to which the experimental groups shows *greater increments in errors* or *smaller decrements in errors* measures the degree of retroactive inhibition. In Table II the average increments and decrements in error scores between the last trial of original learning and successive trials of relearning are listed for both groups. A positive score indicates an increment in errors as compared with the last trial of original learning; a negative score indicates a decrement in errors. Both groups at first show an increment in errors and then a decrement in errors as relearning proceeds. The experimental group, however, consistently shows greater increments in error scores and smaller decrements. The differences between the two groups with respect to increments and decrements in errors and the reliability of these differences are also indicated in Table II. These differences give us a measure of

TABLE II

AVERAGE INCREMENTS AND DECREMENTS* IN ERROR SCORES BETWEEN LAST TRIAL OF ORIGINAL LEARNING AND SUCCESSIVE TRIALS OF RELEARNING FOR EXPERIMENTAL GROUP AND CONTROL GROUP. (Entries include Difference Between the Groups and Tests of Significance of the Differences.)

Relearning Trial No.	Average Increments and Decrements in Error Scores				
	Expt'l	Control	Difference	*t*	*P*
1	6.64	2.86	3.78	5.82	.01
2	1.73	.09	1.64	2.16	.02–.05
3	.18	−1.41	1.59	1.63	.10
4	−1.41	−3.18	1.77	2.57	.01
5	−2.59	−3.55	.96	1.55	.10–.20
6	−3.64	−4.64	1.00	1.25	.20

* Positive entries indicate an increment in errors; negative entires a decrement in errors.

retroactive inhibition *in terms of error scores*. It may be noted that the differences between the experimental group and the control group are significant at a high level of confidence for the first four relearning trials. The differences begin to become less reliable on the fifth and sixth relearning trials.

When differences in error scores have been taken into account, does a comparison of reaction times for correct responses yield any further evidence for retroactive inhibition? Table III shows the

TABLE III

AVERAGE REACTION TIME FOR CORRECT RESPONSES OF EXPERIMENTAL GROUP AND CONTROL GROUP ON ALL TRIALS. (Entries Include Differences Between the Groups and Tests of Significance of the Differences.)

	Average Reaction Time (Sec.)										
	Original Learning					Inter-polation	Relearning				
Trial No.	Expt'l	Control	Diff.	*t*	*P*	Expt'l	Expt'l	Control	Diff.	*t*	*P*
1	1.60	1.69	−.09	2.15	.05–.02	1.58	1.76	1.29	.47	13.8	<.01
2	1.39	1.48	−.09	1.61	.20–.10	1.36	1.71	1.17	.54	13.2	<.01
3	1.26	1.34	−.08	2.00	.05	1.23	1.38	1.09	.29	8.3	<.01
4	1.18	1.22	−.04	1.14	.30–.20	1.15	1.27	1.06	.21	6.8	<.01
5	1.10	1.15	−.05	1.47	.20–.10	1.08	1.18	1.02	.16	4.7	<.01
6							1.14	.99	.15	4.6	<.01

average reaction time for *correct responses* of both groups on all trials. The differences between the experimental group and the control group and the reliability of these differences are also indicated. Under each condition, as learning progresses the average reaction time decreases for both groups. On the original learning trials, the

average response time of the experimental group tends to be shorter than that of the control group although the differences are not very reliable and satisfy the five percent level of confidence only on the first and third trials. On the other hand, *the experimental group consistently shows longer response times on all relearning trials.* On each of the relearning trials the difference between the two groups is significant at a level of confidence of less than one percent.

To gauge the extent of retroactive inhibition as measured by reaction time for correct responses, it is again necessary to compare the performance of the two groups on the last trial of the original learning with their performance on successive trials of relearning. Residual retroaction, i.e., interference which did not manifest itself in retention loss, is measured by the extent to which the experimental group shows greater increments in reaction time or smaller decrements in reaction time than the control group. In Table IV the

TABLE IV

AVERAGE INCREMENTS AND DECREMENTS* IN REACTION TIME BETWEEN LAST TRIAL OF ORIGINAL LEARNING AND SUCCESSIVE TRIALS OF RELEARNING FOR EXPERIMENTAL GROUP AND CONTROL GROUP. (Entries Include Difference Between the Groups and Tests of Significance of the Differences.)

Relearning Trial No.	Average Increments and Decrements in Reaction Time (Sec.)				
	Expt'l	Control	Difference	t	P
1	.66	.14	.52	26.5	<.01
2	.61	.02	.59	21.1	<.01
3	.28	−.06	.34	11.3	<.01
4	.17	−.09	.26	9.0	<.01
5	.08	−.13	.21	6.4	<.01
6	.04	−.16	.20	5.9	<.01

* Positive entires indicate an increment in reaction time; negative entries a decrement in reaction time.

average differences in reaction time between the last trial of the original learning and successive trials of relearning are shown for both groups. Positive entries indicate an increment in reaction time, and negative entries indicate decrements in reaction time. Both groups show increments in reaction time when the last trial of original learning is compared with the first trial of relearning, but reaction times decrease as relearning proceeds. The experimental group shows reliably and consistently larger increments and smaller decrements in reaction time than the control group. The differences between the two groups with respect to increments and decrements in reaction time are listed in Table IV. *On all trials* the differences between the two groups are significant at a level of less than one percent. Thus, even after differences in retention have been taken

into account, reaction time measures differentiate significantly between the two groups throughout the relearning.

The analysis of the results thus shows that there are two possible measures of retroactive inhibition: (1) in terms of retention scores, such as error scores, and (2) in terms of reaction times for correct responses. These two measures are supplementary, since reaction time provides a measure of *residual* retroaction, i.e., such interference as remains when only correctly remembered items are considered. How are these two types of measures related to each other? Fig. 1 gives a graphic picture of the results. Average reaction times for correct responses and average error scores are plotted against trial

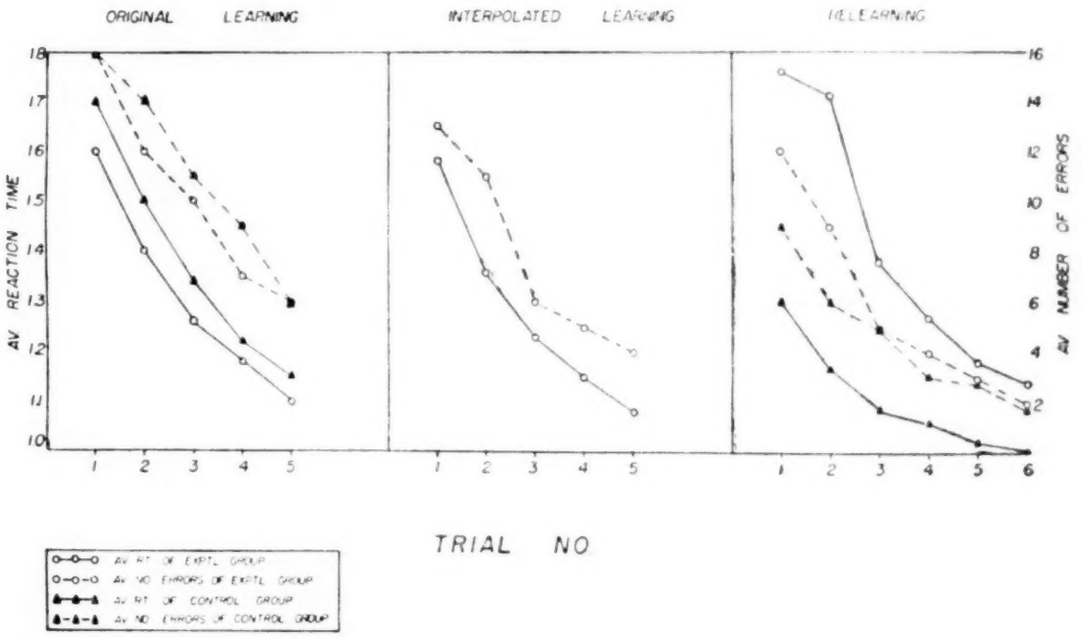

FIG. 1. Average error scores and average reaction times for correct responses of experimental group and control group on all trials

number. Under all conditions of the experiment and for both groups, error scores and reaction times decrease steadily as learning progresses. To that extent, i.e., as far as the general trend of the measures is concerned, they show perfect co-variation. But the additional question remains whether those individuals who show the greatest increments in error scores on the relearning trials also show the greatest increments in reaction time. Rank-order correlation coefficients were, therefore, computed between amount of retention loss (increments in errors) and reaction times on each of the relearning trials. The results are presented in Table V.

For both groups, the correlations between amount of retention loss and length of reaction time for correct responses tend to be low. In terms of rigorous statistical criteria, none of the correlations pre-

TABLE V

RANK-ORDER CORRELATION COEFFICIENTS BETWEEN AMOUNT OF RETENTION LOSS AND LENGTH OF REACTION TIME ON SUCCESSIVE TRIALS OF RELEARNING FOR EXPERIMENTAL GROUP AND CONTROL GROUP

	Rho	
Trial No.	Expt'l	Control
1	.23	.45
2	.23	.11
3	−.18	.08
4	−.16	−.12
5	−.17	−.24
6	−.32	−.24

sented in Table V may be considered significant.[1] In the case of both groups, however, the column of correlation coefficients shows an unmistakable trend toward increasing negativity as a function of trial number. As relearning proceeds, high retention losses show a tendency to be accompanied by short average reaction times for correct responses. It is important to note that the general pattern of correlations between retention loss and reaction time is closely similar for the experimental group and the control group.

DISCUSSION OF RESULTS

The results of the experiment have confirmed the hypothesis that measurement in terms of retention scores does not gauge the full extent of retroactive inhibition. When only *correct* responses are considered, measures of reaction time significantly differentiate between the experimental group and the control group. Differences in reaction time thus measure residual retroactive inhibition after differences in retention scores have been taken into account.

The fact that there are alternative (though supplementary) ways of measuring the effect of an interpolated activity on the retention of an initial learning task raises a general problem of *definition:* in terms of what operations of measurement should *retroactive inhibition* be defined? If one chooses to define it in terms of sheer retention loss, then measures of reaction time for correct responses would, strictly speaking, be irrelevant. Similarly, if a definition were made in terms of reaction times, measures of retention would cease to be relevant. If no definition in terms of specific measurements is made, the term loses its operational referent and the results of different experiments cease to be comparable. The results of this experiment lead to the conclusion that retroactive inhibition should be defined as referring

[1] To evaluate the rank-order correlation coefficients a formula suggested by Lindquist (3) was used: $\sigma_\rho = 1.05\ (1 - \rho^2)/\sqrt{N}$. To be statistically significant rho would have to be three times as large as its standard error. None of the coefficients meets this criterion.

to *all measurable aspects of interference with memory* which may be ascribed to the effects of an interpolated activity. Two such measurable aspects of interference are (1) decrements in retention scores and (2) reaction time for correct responses. A definition in terms of these two measures is not final. Other measurable aspects of interference may be added. Additional measures should be included if they differentiate significantly between the experimental group and the control group after other measures of retroaction have been taken into account. Since retroactive inhibition is defined in terms of the operations by which it is measured, the greater the number of independent and supplementary measurements, the more complete the definition of the process will become.

The question of *correlation between measures* remains to be considered. As Table V shows, the two measures used in this experiment are not significantly correlated. There is, however, a certain trend toward increasing negativity of correlation as relearning progresses. The lack of correlation may be interpreted to mean that the two measures reflect essentially independent processes. It has been argued previously that the total retroactive effects, as they are measured in an experiment, may be the resultant of different processes of interference acting simultaneously. Thus, Melton (**6**) has suggested a two-factor theory of retroactive inhibition. The two factors are (a) the unlearning of the original activity during the interpolation and (b) competition between original and interpolated responses during recall. If these two factors are independent of each other, measures of their operation should not be expected to be correlated. It is possible that retention loss (error scores) reflects primarily the effects of unlearning, whereas reaction times may depend largely on the competition between responses. Some doubt is thrown on this interpretation, however, by the fact that the pattern of correlations between the two measures is very similar in the case of both groups. The control group was not exposed to an interpolated learning task, and competition between responses can, therefore, not be expected to occur with these *S*s.

The data of this experiment do not permit us to speculate further about the nature of the events which result in retroaction. The lack of correlation between two independent measures of the effect suggests, however, that retroactive inhibition should not be regarded as a unitary process. Since two independent and uncorrelated measures of retroaction can be obtained, we must assume, at least for the time being, that these two measures reflect two distinct processes. The results of a previous investigation (**7**) in which retroactive inhibition was studied as a function of the temporal point of interpolation supported the same conclusion. The empirical functions could best be

explained by postualting two distinct processes following different courses in time.

The trend toward increasing negativity of correlation shown in Table V [2] seems to indicate that the two processes begin to take different courses in time as relearning proceeds. Those individuals whose error scores improve most tend to show the highest average reaction times on successive relearning trials. Conversely, those individuals who have high error scores show short reaction times relative to other members of the group. It seems that improvement alone one dimension may occur at the expense of improvement along another. *S*s' relearning may, to a certain extent, be selective. A *limited* number of items may be relearned and overlearned, yielding shorter and shorter reaction times. Other items, on the other hand, may remain unavailable for recall on several relearning trials. Thus the selectivity of forgetting and relearning may account for the divergence of temporal trends in the two independent measures of retroactive inhibition obtained in this experiment.

An alternative explanation of the negative trend of correlations is possible. The divergence in the two measures of retroaction may be due to differences of *attitude* among the *S*s. In the later relearning trials some *S*s may be more cautious than others and take time for deliberation rather than give wrong responses. Other *S*s may emphasize speed at the expense of accuracy, thus giving quick, correct responses but also more false responses than the more cautious *S*s. This difference in attitude becomes most effective in the later trials because on the earlier trials all reaction times tend to be long, and even the cautious are in no position to deliberate in order to give any response at all in the alloted time.

The results of the experiment do not favor either explanation of the negative correlations. Before deciding in favor of the one or the other, it will be necessary to await further evidence for and against a two-factor theory of retroactive inhibition.

Summary

Retroactive inhibition is usually measured in terms of decrements in recall due to an interpolated activity. Such measures may not be sufficiently sensitive. An *S* may respond correctly, but only after a momentary block or vicarious trial-and-error. It was hypothesized, therefore, that reaction-time might be a sensitive measure of that interference which does not manifest itself in retention loss.

[2] It is true that none of the coefficients can be considered statistically significant. The *trend* toward increasing negativity, however, is unmistakable and makes it highly unlikely that the changes in correlation are random. The only reversal in Table V is from trial 3 to trial 4 of the experimental group. This is not a reversal in sign but merely a small reversal in the magnitude of the coefficient.

The usual paradigm for experiments on retroactive inhibition was followed. The experimental group learned an interpolated task two min. after the original learning, while no interpolation was used with the control group. Both groups were retested 48 hours after the original learning.

The original learning series consisted of 20 pairs of unrelated words. The interpolated learning series consisted of 20 word-pairs whose left-hand members were identical with those in the original learning series while the right-hand members consisted of new unrelated words. The relearning series was identical with the original learning series.

The anticipation method was used. On all trials, reaction time was recorded by means of a voice-key arrangement.

Reaction time for correct responses decreased equally for both groups during the original learning. On *all* relearning trials, the experimental group showed significantly longer reaction times than the control group. During the relearning, reaction times of the control group became significantly shorter than they had been during original learning while the reaction times of the experimental group never reached the original level.

Moderately low correlations were obtained between reaction times and retention loss. As relearning progressed these correlations tended to become negative.

Thus, reaction time provides a measure of retroactive inhibition over and above measurement in terms of retention loss. The low correlations between the two measures suggest that two relatively independent processes may be operative.

(Manuscript received May 31, 1946)

References

1. Ach, N. *Ueber den Willensakt und das Temperament.* Leipzig, 1910.
2. Lewin, K. Das Problem der Willensmessung und das Grundgesetz der Assoziation. *Psychol. Forsch.*, 1921, 1, 191–302 and *ibid.*, 1922, 2, 65–140.
3. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton-Mifflin, 1940.
4. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green and Co., 1945.
5. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.
6. Melton, A. W., & VonLackum, W. J. Retroactive and proactive inhibition in retention: evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, **54**, 157–173.
7. Postman, L., & Alper, T. G. Retroactive inhibition as a function of temporal point of interpolation. *Amer. J. Psychol.*, 1946, **59**, 439–449.
8. Woodworth, R. S. *Experimental psychology.* New York: Henry Holt and Co., 1939.

A STUDY OF MUSCLE ACTION POTENTIALS DURING THE ATTEMPTED SOLUTION BY CHILDREN OF PROBLEMS OF INCREASING DIFFICULTY *

BY W. A. SHAW AND L. H. KLINE

University of Pennsylvania

Studies by Davis (1) and by Hadley (2) have shown that an increase in muscle action potential accompanies tasks of increasing difficulty as measured by the percentage of failure. In these experiments adult *S*s attempted to solve graded arithmetic problems without the use of pencil and paper. It was found in both experiments that action potentials increased with the increasing difficulty of the problems. The immediate problem to be considered in this paper deals with the study of muscle action potentials during the attempted solution of problems of increasing difficulty by children.

Apparatus

The apparatus used in this experiment is the same as that used by Shaw (3). Briefly, it consists of a resistance-capacitance coupled amplifier, a five-in. cathode ray oscillograph, a beat-frequency oscillator and a camera.

Two small rooms were constructed, one to house the apparatus and the other to screen the subject. The former measured 5′6″ long, 5′ wide and 7′ high, and was constructed of a wood frame covered with fiber-wood. The latter measured 6′ × 6′ × 7′ and was constructed of a wood frame covered with No. 24 galvanized iron. This room, together with the cabinet in which the amplifier was mounted, were grounded to the water pipes. Both rooms were painted black inside.

The *S*'s room contained a special chair equipped with a headrest to hold the *S*'s head in a stable position. It also contained a table on which was mounted a stand. This stand supported a small fixture for holding 5 × 7 in. cards on which the problems were printed.

Shielded leads were used to connect the *S* to the amplifying system. These leads were terminated in surface electrodes similar to those used by Davis (1).

Communication was carried on between the two rooms by means of a mechanical signalling system.

Procedure

Ten male *S*s were used, whose ages ranged from 10 years and 4 months to 13 years and 9 months. Their IQs ranged from 74 to 124. All of the children were known to *E* No. 2, who took them to the experimental room a few times before the tests were begun. *E* No. 2 discussed with them the various items which they found interesting in the experimental room. During these orienting visits *E* No. 2 also attempted to motivate the children by telling them of the 'new' tests he had in store for them.

As soon as *S* arrived at the laboratory he was seated in the shielded room and the electrodes were attached to the lower right arm. The active electode was placed on the outer surface of the

* The authors wish to express their thanks to the University of Pennsylvania Chapter of Sigma Xi for a grant covering the cost of film used in this study.

upper forearm; the inactive electrode was placed on top of a felt pad moistened with physiological salt solution and the pad was fastened on the inside of the elbow. *E* No. 2 then gave *S* instructions to rest. Since it was necessary to have an observer in the shielded room with the *S*, *E* No. 2 seated himself at the right side of *S* and as far from him as the room would permit. His task was to observe the course of *S*'s activity, to aid *S* to relax whenever necessary, to caution him continually against movement and change of position, and to administer the test and record the results.

S was usually allowed to rest from 15 to 20 min. before the experiment began. If the rest level had not fallen to 10 microvolts or below by this time, he was given an extra 10 min. of rest. At the end of this time, if rest was still unsatisfactory, the *S* was permitted to leave. During this resting period the course of *S*'s activity was observed on the oscillograph and suggestions for relaxation were offered by *E* No. 1 who controlled this apparatus. No records of the initial rest were taken until the criterion of 10 microvolts or below obtained. As soon as *S* reached this level, his activity was sampled, and at the end of this time *E* No. 2 presented him with the sample problems for inspection and explanation. Standard instructions typed on a card were read after the sample problems were shown. These instructions were:

1. Sit as still as you can. Do not move your hands, legs, or head.
2. I shall ask you to find the answers to some problems in arithmetic without using pencil and paper. The problems are in addition (adding up), subtraction (take away), multiplication (times), and division (goes into).
3. I will place a card with the problem on it here (points to stand), and you will begin to work on the problem at once.
4. Give your answer to me as soon as you think you have it. If you cannot find the answer after you have tried, then tell me right away. Do not guess. There is no time limit.
5. Remember, sit as still as you can, and do not move your hands, legs, or head.

The *S* was then asked if he understood what he was supposed to do. If he did not understand, the instructions were again read. If he indicated that he did understand, No. 1 and No. 4 were repeated.

The problems and their respective PE values were selected from Woody's Series 'A' Arithmetic Scales (4). Five problems each in addition, subtraction, multiplication and division were selected. The five groups had average PE values of 1.40, 2.61, 3.63, 4.79 and 6.10. Our purpose was to secure a range of problems the easiest of which could be solved by all the *S*s and the most difficult by none. This aim was not quite realized, but the selection of problems in this manner was entirely satisfactory.

These 20 problems were each typewritten on a plain 3 by 5 in. index card. The cards were shuffled before being presented to the *S* in an attempt to minimize whatever effects might follow upon the presentation of a constant order. Each card was placed on the small, well-illuminated stand in front of the *S* who then attempted to solve it without the aid of pencil and paper. As each card was presented the *S* was told, "This is an addition problem," etc. As soon as the *S* was presented with the problem, *E* No. 2 signaled *E* No. 1, who began taking a record of the activity shown on the oscillograph. The time *S* worked on each problem was recorded by *E* No. 2 who again signaled *E* No. 1 at the conclusion of the work. After *S* had offered a solution, he was instructed to rest. This rest lasted long enough for him to approximate his initial rest level very closely. When this obtained, a sample of rest was taken and *E* No. 1 signaled to *E* No. 2 to proceed.

E No. 1 sampled the course of *S*'s activity during work by running the camera for five sec. and pausing five sec. begining with the initial five sec. of activity. The same procedure could not be followed during rest, consequently records of *S*'s activity during rest were taken during two five-sec. intervals separated from each other by a like interval after *S* had reached the criterion of rest. The intervals were timed by a stopwatch. The camera ran at the rate of eight frames per sec.

In measuring the film each frame was measured by projecting it onto a ruled screen. This enlarged the record 50 times and permitted a high degree of accuracy. By this method it was possible to read deflections of less than a microvolt. The sensitivity of the amplifier was adjusted to approximately one mm. (one direction) per microvolt input. The use of greater sensitivity tended to produce many records in which the amplitude of the recording trace was greater than the film could record. The level of amplification was the same for all *S*s and was checked after every fifth problem throughout the course of the experiment by the use of the oscillator.

The method of treating the measurements consisted of averaging the deflections which occurred during the solution of each problem of the five sec. yielding the greatest amplitude. Other treatments were tried such, for example, as averaging all of the deflections for each work period, and again averaging the deflections for the highest second in each work period. Both treatments yielded the same trends as the adopted treatment. However, averaging all deflections in each work period attenuates the values for each work period, since the period between the signal to begin work and the actual beginning of the work is included. Averaging the deflections for the highest sec. in each work period increases the variability of the values. Averaging the deflections for the five sec. yielding the greatest amplitude represents a compromise such that maximum average deflections are obtained consistent with a reduction in variability. These averages were then converted into microvolts and this value was taken as an index of the activity of the *S* during the work or rest period.

Another way of expressing the results is the work-rest ratio suggested by Davis (1). To get these ratios the activity value of each of the problems was divided by its respective rest value. These rest values are averages of all of the available deflections in the rest period immediately preceding each work period. The ratios are then treated in a manner similar to the absolute values.

Results

1. Grouping the problems and arranging the results: In the procedure it was stated that five groups of problems were selected from Woody's lists (4). Each group contained one problem each in addition, subtraction, multiplication and division. The PE values of the problems in each group were approximately equal. It was not expected that Woody's standardization would determine exactly the degree of difficulty for the present group, since the conditions under which the present group attempted to solve the problems were different. Nevertheless, to a certain degree this standardization of the problems served as an independent criterion of difficulty. This departs from the method used by Davis (1) and by Hadley (2), both of whom assigned difficulty values to their problems on the basis of the errors made by their respective experimental groups.

2. Activity during rest: In Table I are the average values of the activity in the rest periods immediately preceding each problem, arranged according to the level of difficulty of the problems which they preceded. These values are somewhat higher than the initial and

TABLE I

Average Activity Level During the Rest Periods.

Values are expressed in microvolts

Problem Group	1	2	3	4	5
N (rest periods)	39	39	40	40	39
Ave.	9.57	9.30	9.32	9.42	9.72
SD	2.83	2.32	2.60	2.89	3.34
V	29.57	24.94	29.89	30.68	34.36
Ave. above 100 IQ (Group A)	8.16	8.41	8.46	7.86	8.43
Ave. below 100 IQ (Group B)	10.98	10.22	10.29	11.00	11.09

final rest values found in the experiments of Davis and of Hadley mentioned above. The former gives values for the arm of 5.9 and 5.7 microvolts for the initial and final rests respectively; the latter lists values of 6.74 and 8.42 microvolts for the initial and final rests. The somewhat higher rest levels found in the present experiment may be due to the inability of the subjects to grasp fully the idea of relaxing. The maximum difference in variability (V) from problem group to problem group is approximately 38 percent. The variability of these rest samples compares favorably with values computed from Davis's results (32.8), but they are somewhat lower than those computed from the results of Hadley (35.6 to 54.2).

For the purpose of further analyzing the data, the 10 *S*s were divided into two groups. In Group A were placed the five *S*s whose IQs were greater than 100. The five *S*s whose IQs were below 100 were placed in Group B. When the data are broken down according to whether the rest samples belong to individuals having IQs above or below 100, it can be seen that Group A has a lower average activity level than Group B.

Certain factors operated to delay the rest level from returning to the initial level once the experiment was begun. First, the children tended to rework the problems. This held true more for the higher intelligence group. Secondly, they tended to consider the rest periods as intervals during which they could relieve themselves of the task of sitting quietly. This tendency to move about during the rest periods was more characteristic of the low intelligence group.

It was noticed that there were rather wide individual differences among the *S*s used in their ability to relax, using approximately 10 microvolts as a criterion. Some *S*s relax quickly upon being asked to do so. Others seem to have little control over their tenseness and succeed in relaxing only after a protracted period of time and considerable training. Three *S*s whose results are not here reported were rejected because of their inability to rest.

3. The treatment of the samples of activity during the work period: The results of the *S*s have been treated first as a whole and later according to whether they belong to *S*s having an IQ above or below 100. When the work samples are grouped according to the difficulty values of the problems, an increase in activity occurs with each succeeding step of difficulty for the group as a whole, except for the problems in group 5. Referring to the averages in Table V, it may be seen that for the group as a whole there are more errors on the problems in group 4 than on the problems in group 5. This is due apparently to the efforts of the *S*s in the A-group.

4. A comparison of the variability of activity in the rest and work periods: The variability of the work periods shown in Table II is, on

TABLE II

AVERAGE OF THE MEASURES OF THE ACTIVITY LEVEL DURING THE WORK PERIODS.
Values expressed in microvolts

Problem Group	1	2	3	4	5
N (problems)	37	39	38	37	39
Ave.	12.75	13.51	14.89	18.59	17.21
SD	5.29	7.44	5.11	8.42	9.51
V	41.49	55.07	34.32	45.29	55.26
Ave. above 100 IQ (Group A)	9.81	10.10	11.52	15.77	12.05
Ave. below 100 IQ (Group B)	15.35	16.90	17.92	21.05	22.12

the average, approximately 46.29, while that for the rest periods is 29.89. The latter is then 64.5 percent as variable as the former. Comparable figures calculated from Hadley's (**2**) results show his rest periods to be on the average 91.5 percent as variable as his work periods. This difference is due to the difference in technique used in the two experiments in handling the rest periods. In general, the variability of the work periods among themselves is greater than that found among the rest periods.

5. Critical ratios between rest and work: The critical ratios between rest and work are significant in all cases. The values presented in Table III have been determined from the averages in

TABLE III

CRITICAL RATIOS BETWEEN REST AND WORK.

Problem Group	1	2	3	4	5
σ_{mean} Rest	.44	.37	.41	.46	.53
σ_{mean} Work	.88	1.21	.85	1.38	1.52
rho (rest and work)	.83	.74	.68	.63	.68
$\sigma_{diff.}$	.57	.97	.53	1.15	1.22
Diff. (rest and work)	3.18	4.21	5.57	9.17	7.49
Diff./$\sigma_{diff.}$	5.58	4.35	10.51	8.01	6.18

Tables I and II. Rho has been used in these calculations instead of *r*. The significance of the difference between problems is not considered of special value since it depends largely upon the relative difficulty of the problems selected. The correlation between the rest samples and the work samples indicates that as the problems become more difficult the activity in the two periods is less related.

6. Work-rest ratios: Another way of presenting the data is to divide the amount of activity during the work period by the amount of activity in the rest period. The results are presented in Table IV. These ratios show the same trends as do the absolute values. In terms of these ratios the A-group has about the same amount of

activity on the problems in groups 1 and 2. Again there is a decrease on the problems in group 5. These problems are easier for the A-group than those in group 4 in terms of the *Ss*' own percentage of failure (Table V).

Work-rest ratios have certain advantages over absolute values, as Davis (1) has pointed out, in ruling out individual differences. There seems to be some evidence in Table IV which indicates that

TABLE IV

RATIOS OF WORK AND REST. These values were computed from values in Tables I and II

Problem Group	1	2	3	4	5
N (problems)	37	39	38	37	39
Ave.	133	145	160	197	177
Ave. above 100 IQ (Group A)	120	120	136	200	143
Ave. below 100 IQ (Group B)	140	165	174	192	200

differences due to the effect of factors which may accompany differences in intelligence cannot be equated in this manner. For example, the A and B groups made 45 percent failures (Table V) on problems

TABLE V

PERCENTAGE OF FAILURE ON EACH PROBLEM GROUP

Problem Group	1	2	3	4	5
Ave. for group	24.3	30.8	55.3	94.6	92.3
Ave. above 100 IQ (Group A)	11.1	15.8	44.5	94.1	84.2
Ave. below 100 IQ (Group B)	36.9	45.0	65.0	95.0	100.0

in groups 3 and 2, respectively. In terms of the percentages of failure these two problems have the same difficulty value. They do not, however, have the same activity value for the two groups. On the other hand, the A and B groups have the same amount of activity in terms of work-rest ratios on the problems in groups 4 and 5. In the former case the A-group exhibits less activity on the problems of the same difficulty for the B-group; in the latter case the A-group has fewer errors with the same amount of activity. The A-group 'works' less hard on problems of the same degree of difficulty for the B-group, but it 'works' harder as the difficulty increases than does the B-group. The data, however are too few to state definite conclusions.

7. The influence of intelligence level or of factors accompanying various levels of intelligence upon activity during the attempted

solution of arithmetic problems: The values of work in Tables II and IV show that there is a decrease in activity on the problems in group 5 for the Ss as a whole. This decrease in activity seems to be due to the fact that the Ss in the A-group did not find these problems as difficult, in terms of percent failure, as did the B-group. Reference to Table V, in which is given the percentage of failure on each group of problems, shows that the A-group made 10 percent fewer failures during the attempted solution of the problems in column 5 than in column 4. The problems in column 4 are more difficult for the A-group when their own percentage of failure is taken as the criterion. Standardization of the problems on an independent group did not hold good for the A-group throughout the whole range.

Group B not only makes more errors on each group of problems than does Group A, but has higher rest and work activity levels. Group B also has greater relative activity over the problems in the first three classifications. There is some doubt whether Group B's relative activity level can be compared to Group A's relative activity level on problems in classifications 4 and 5, because of the fact that standardizing the problems beforehand did not determine their difficulty for the A-group.

8. The time of attempted solution and its relation to activity and difficulty: The time taken for the *S* to offer a solution or to give up has not been made use of in previous studies of this nature. The time of attempted solution of the problems for the group as a whole (Table VI) increases over the first four steps of difficulty and then

TABLE VI

AVERAGE TIME IN SEC. FOR THE ATTEMPTED SOLUTION OF THE PROBLEMS ARRANGED ACCORDING TO PROBLEM GROUP

Problem Group.....	1	2	3	4	5
Ave. sec. per prob.....	10.2	13.2	23.6	36.3	35.2
Ave above 100 IQ (Group A).....	6.5	11.9	27.5	43.5	38.9
Ave. below 100 IQ (Group B).....	13.9	14.5	19.8	31.4	31.4

seems to level off. But the times taken for the A and B groups are very different. The time for the A-group increases rapidly from problem group to problem group, with the exception of the problems in group 5. The A-group takes less time on the easier problems and more on the harder ones when compared to the B-group. One might expect the B-group to take more time than the A-group, even on the harder problems. Among other possible factors there are two important ones which prevent this expectation from being realized, namely: (1) The tendency for the B-group to report answers before

reaching a thorough understanding of the problems; (2) the marked tendency for the B-group to give up.

9. The use of the time taken for attempted solution; of success or failure to report the correct answer; of whether the *S* reported an answer or gave up; in conjunction with observations of the *S*s at work for assigning difficulty ranks to problems for each *S*: Inspection of the data in Tables V and VI and of the *S*s' reports suggests another mode of arranging the problems to conform more closely with the *S*s' performance upon each problem. The following observations from the *S*s' reports were drawn upon for ranking the problems according to difficulty. Some problems were more difficult than others because (1) it took more time to solve some problems than it did to solve others; (2) some problems could be solved while others could not; (3) on some problems the *S* had to 'work harder' in order to reach a solution than he did on others; (4) the *S*s reported 'working harder' on some problems than on others when neither was solved; (5) some problems were 'harder' than others because the *S* "didn't know how to do it."

For each *S* the 20 problems were ranked more or less arbitrarily in the following manner: (1) problems which were solved were ranked according to the amount of time taken in order to reach the solution; (2) problems for which wrong answers were offered were ranked according to the amount of time spent upon them; (3) of two problems, one of which was solved and a wrong answer offered for the other, the latter was considered the most difficult unless the former took twice or more than twice the time to solve; (4) in all cases, those problems on which the *S* gave up were considered the most difficult. Of two problems on both of which the *S* gave up, that which took the least time was considered the most difficult. This procedure was purely exploratory.

10. Average time in sec. for the attempted solution of the problems when arranged according to the method of ranking suggested above: When the problems are ranked according to the above rules, the time for each resulting problem group may be seen in Table VII

TABLE VII

AVERAGE TIME IN SECONDS FOR THE ATTEMPTED SOLUTION OF THE PROBLEMS WHEN RANKED

Rank	1	2	3	4	5
Ave. sec. per prob.	6.9	12.4	20.7	34.2	46.1
Ave. above 100 IQ (Group A)	6.2	11.4	15.2	35.0	59.6
Ave. below 100 IQ (Group B)	7.5	13.3	25.6	33.4	33.8

to differ from the first arrangement of the problems in several ways. The Ss as a whole worked longer with each succeeding rank. The range of times is widened. The A-group shows little change over the first two ranks. On ranks 3 and 4 these Ss show smaller average times than in the first arrangement (Table VI). The problems in rank 5 have a greater average time. The B-group shows that the average time on the problems in rank 1 was not greatly different than that for the A-group. For the B-broup there is little change in the average time for the problems in rank 2. The problems in rank 3 show a greater average time. The problems in ranks 4 and 5 show the same relation to each other as on the first arrangement. On the whole then, ranking the problems according to the above rules has in effect arranged the times for attempted solution in the order of increasing times.

11. Average number of errors for each successive rank: When the problems are ranked according to the preceding rules, the average number of errors per rank increases for both the A and B groups over the whole range. In Table VIII it may be seen that the greatest

TABLE VIII

PERCENTAGE OF FAILURE ON EACH RANK

Rank	1	2	3	4	5
Ave. for group	21.1	33.3	60.5	83.8	100.0
Ave. above 100 IQ (Group A)	5.6	15.7	55.6	72.2	100.0
Ave. below 100 IQ (Group B)	35.0	50.0	65.0	94.7	100.0

change in the present grouping compared with the first grouping takes place in the average number of errors per rank for the A-group. Ranking the problems according to the above rules results in both the A and B groups failing all of the problems in the fifth rank. When the problems are arranged in this manner, there result five groups of problems on which the Ss took more time and on which they made more errors with each successive group. The question then arises as to the nature of the Ss' activity during the attempted solution of the problems thus arranged.

12. Average activity during the rest period immediately preceding each problem arranged according to rank: When the amount of activity during the rest intervals preceding each problem are arranged to follow in the same order as the problems, very little change takes place over the original grouping of the rest samples. There is a tendency for whatever trends that appear in the first method of grouping the problems (Table I) to be emphasized, otherwise the general picture of the rest levels preceding each problem remains unchanged.

TABLE IX

AVERAGE ACTIVITY LEVELS DURING THE REST PERIODS ARRANGED ACCORDING TO THE RANK OF THE PROBLEM WHICH IT PRECEDES. Values expressed in microvolts

Rank	1	2	3	4	5
N (rest periods)	39	40	39	40	39
Ave.	9.67	9.03	9.59	9.57	9.91
SD	2.76	2.25	2.60	3.13	3.84
V	28.54	24.92	27.11	32.71	38.75
Ave. above 100 IQ (Group A)	8.43	8.08	8.52	8.10	8.29
Ave. below 100 IQ (Group B)	10.95	9.99	10.61	11.04	11.53

13. Activity during the work periods when the problems are ranked: Activity during the work periods increases with each succeeding rank throughout the entire range, not only for the group as a whole, but also for both the A and B groups. The variability remains about the same from rank to rank as in the previous arrangement of the problems. The same trends appear in the average values in Table X as in the previous table, with the exception of the most

TABLE X

AVERAGE ACTIVITY LEVELS DURING THE WORK PERIODS WHEN THE PROBLEMS ARE RANKED. Values expressed in microvolts

Ranks	1	2	3	4	5
N (problems)	38	38	38	37	39
Ave.	12.32	13.49	15.62	17.17	19.92
SD	4.61	7.47	7.52	8.46	9.43
V	37.41	55.38	48.14	49.27	47.34
Ave. above 100 IQ (Group A)	9.63	10.49	11.44	13.84	15.45
Ave. below 100 IQ (Group B)	14.47	16.34	19.38	20.33	24.17

difficult problems. Ranking the problems has relatively more effect upon the A-group than upon the B-group. In both groups the effect is more pronounced over the last two ranks.

14. Activity during the work period expressed in work-rest ratios when the problems are ranked: The tendencies brought out in Table X are also present when the data are expressed in the work-rest ratios shown in Table XI. The advantage of using such ratios has been discussed earlier. From inspection of these three tables there is little doubt but what activity increases with each succeeding rank. When the problems are ranked according to the rules previously stated in such a manner that the time of attempted solution and the number of errors increase, muscle action potentials also increase. It is believed that closer attention to the responses of the *S* during attempted solu-

TABLE XI

RATIOS OF WORK AND REST. The values in the table were computed from the values in Tables IX and X

Ranks	1	2	3	4	5
N (problems)	38	38	38	37	39
Ave.	125	147	162	190	212
Ave. above 100 IQ (Group A)	116	123	136	176	195
Ave. below 100 IQ (Group B)	133	170	186	202	229

tion would result in a more accurate means of ranking the problems and effect a still stronger relationship than that shown. An objection to the method is that at present there is no way of scaling these ranks.

SUMMARY OF THE RESULTS

The study of the relation of muscle action potentials to difficulty in the problem-solving situation is beset with many obstacles. Such obstacles may be multiplied many fold where children are used as *S*s. In spite of the fact that one of the *E*s was known to all of the *S*s, and every precaution was taken to insure the best cooperation obtainable, it was not possible to maintain the close control of attitudes and sets as in the case of adult *S*s. Few children grasp the idea of relaxation readily. They do not understand that small movements are important. They may deny making them, and, if criticized a few times, they tend to become emotional. And, when relaxation finally obtains and work is begun, these small movements may recur. Their function has recently been investigated again by Grinsted (**5**), who believes them to be adjustmental in character. Although of no importance in many experiments, such small movements greatly disturb the recording of action potentials.

Further cause of unwanted disturbance was often traced to special interest in a particular problem. Again, *S*s often recognized in a particular problem a process with which they had previously experienced difficulty. In this case one can expect marked changes in attitude. The *S* may refuse to make an attempt to solve the problem. More often he will simply guess after briefly attempting a solution. The possibility of carrying over a disturbance from one problem to another was minimized by having the *S* rest after each problem.

Interpretation of the results of the preceding section is based upon the assumption that the psychological responses involved in problem solving will vary in either the number, kind, complexity, and intensity, or in all of these factors, as the problems increase or decrease in difficulty value. If we grant this assumption, there are two things

we should like to know. The first is whether muscle activity accompanies the psychological activities present in problem solving when we use muscle action potentials as an index of muscle activity. Secondly, if muscle activity is present, we should then like to know whether there is a concomitant variation in its magnitude which corresponds with the various degrees of problem difficulty.

The results in Table II show that muscle action potentials in significant amounts do accompany the attempted solution of problems. Secondly, as the difficulty value of the problems increases, these results show that muscle action potentials also increase. This relation is more strongly indicated in Table X, where the difficulty value of each of the problems is determined for the present group of *Ss* largely by the use of the length of time it took the *Ss* to arrive at a solution.

The relation between the amount of muscle action potential and the degree of difficulty is still further strengthened by other evidence. For example, in Table V the *Ss* in Group B (below 100 IQ) find the problems as a whole more difficult in terms of the percent failed than do the members of Group A (above 100 IQ). The work-rest ratios in Table IV show that the members of Group B are also more active during the times they worked on the problems. This finding is further substantiated by the results in Tables VIII and XI.

Because of the small number of cases used and because it is not known exactly what aspects of behavior vary with the increasing difficulty value of the problems, it is not believed practical at this time to attempt the development of a more precise relationship between muscle action potentials and the degree of difficulty in problem solving. It is more important to show that if we can assume that some aspect (or aspects) of behavior in the problem-solving situation varies with increasing difficulty, we can establish an independent variable. If, then, we find, as the present results seem to indicate, that the dependent variable (action potentials) changes with corresponding changes in our independent variable, we must conclude that the two are related.

Conclusions

1. When the children attempted the solution of a pre-graded series of problems without the use of pencil and paper, muscle action-potentials in their lower right arms increased as the problems increased in difficulty.

2. When the problems were arranged to take account of the time, and of the number and kind of errors made in the solution of the problems, the tendency for muscle action-potentials to increase was more marked.

3. Within the limits of the *S*s used, the group having the greatest intelligence showed the least activity both during rest and during the solving of the problems.

(Manuscript received March 18, 1946)

References

1. Davis, R. C. The relation of muscle action potentials to difficulty and frustration. *J. exp. Psychol.*, 1938, **23**, 141–158.
2. Hadley, J. M. Some relationships between electrical signs of central and peripheral activity: II. During 'mental work.' *J. exp. Psychol.*, 1941, **28**, 53–62.
3. Shaw, W. A. The relation of muscular action potentials to imaginal weight lifting. *Arch. Psychol.*, N. Y., 1940, No. 247, p. 16.
4. Woody, C. Measurement of some achievements in arithmetic. *Columbia Univ. Contr. Educ.*, 1920, No. 80.
5. Grinsted, A. D. Bodily movement as related to problem solving. *J. exp. Psychol.*, 1941, **29**, 370–379.

REMINISCENCE FOLLOWING EXPERIMENTALLY INDUCED FAILURE

BY DAVID C. McCLELLAND AND F. S. APICELLA[1]

Wesleyan University

On the basis of a series of experiments (**5, 6, 7**) the senior author arrived at the hypothesis that reminiscence is caused by the removal of inhibition and not by a differential rate of forgetting as has been customarily assumed (**1, 9**). The arguments on which the hypothesis was based were somewhat inconclusive, however, and in most instances rather involved. In no case was there a datum which could be explained only by the removal of inhibition hypothesis and not by an extension of the differential rate hypothesis. And neither theory satisfactorily explained all the data subsumed under the heading of reminiscence.

Part of the difficulty underlying this theoretical uncertainty is methodological. The experiments purporting to demonstrate the more rapid loss of wrong associations used rote learning materials which did not permit the easy identification of learned wrong responses nor the measurement of their curve of forgetting. The senior author's attempt (**5, 6**) to get around this difficulty by using discrimination learning did not succeed entirely, and ran into the added difficulty of trying to explain reminiscence in terms of performance or recitation conflicts, the occurrence of which was not any better controlled than the occurrence of the wrong associations in ordinary rote-learning experiments.

The next step in the development of the removal of inhibition hypothesis is not to attempt to infer afterwards what conflicts must have been present in a given learning task, but to introduce controllable conflicts into the learning so that they may be removed by rest to test the hypothesis that such removal should produce reminiscence. Such an experiment, although it will not demonstrate that the differential rate of forgetting theory cannot apply in some instances, will at least show that in some instances removal of inhibition *is* what causes reminiscence. It will also serve to define by a set of operations what is meant by 'inhibition,' a term which has been used throughout

[1] The senior author is responsible for the experimental design and the present report, the junior author for the execution of the experiment.

this series of papers in the sense of performance decrement caused by conflicts of any sort.[2]

The experiment reported here is designed to take this next step in the corroboration of the removal of inhibition hypothesis. It starts with the finding reported by Sears (**10**) that *S*s who are made to fail continuously will perform consistently less well than others who are allowed to succeed. The poorer performance is interpreted here to mean that the *S*s suffered from conflicts caused by anticipations of failure. If this interpretation is correct, removal of these anticipations of failure by rest in which success is experienced should result in reduced conflict and improved performance or reminiscence. So one group of *S*s was forced to fail in the performance of a task, and then was rested with a short period of success on another task, to see whether on resuming the first task they would perform better for having had some of the failure-induced conflicts removed. Two control groups were added to exclude other possible explanations of the results. The first consisted of *S*s who were not made to fail but who, after reaching the same learning criterion, were given the rest with success. It was needed to see whether whatever changes occurred after rest would appear irrespective of treatment in original learning. The second consisted of *S*s who were made to fail but who were not given a rest before relearning. It was needed to see whether changes which occurred after rest might occur anyway if no rest had been introduced.

Procedure

The task used was sorting an ordinary pack of playing cards into the four different suits according to the principle that each card turned up must be placed on the suit symbol of the card just previously turned up. The time it took the *S* to sort through the cards was taken with a large Eastman timer which could be shown to the *S* occasionally to verify a score the experimenter reported.

The aim of the experiment as explained to each *S* was to sort through the pack of cards as quickly as possible, correcting each mistake, until he could complete a trial in 45 sec. After a few practice trials to familiarize him with the technique he was asked to state how many seconds he expected to take to sort the pack. With this the experiment proper began. After every trial he was told the number of seconds it took him and was asked what he expected to do it in next time. What score was reported depended somewhat on what he really did (his real score) and somewhat on what he had expected to do (level of aspiration). The false or reported scores were manipulated within the limits of these two determinants so that he failed on every two out of three trials and just reached his level of aspiration on the third. The object of the technique was to upset each *S* as much as possible, rather than to put each one through a standard sequence of successes and failures (which was relatively impossible anyway because of the variability of stated levels of aspiration). Therefore, the sequence of reported failures was varied to suit the individual case. Furthermore, the *E* complained continuously about the *S*s' failures, frequently swearing at them in his exasperation (cf. **8**, p. 378, for a full description of the type of comment). The net effect of this treatment was to make many of the subjects very angry and their remarks (cf. **8**, pp. 381–382) leave little doubt that they were experiencing genuine failure.

[2] Sources of such conflicts have been mentioned in previous articles and are considered in some detail elsewhere (**4**).

The second task which was used to create an experience of success in some *Ss* was the standard Koerth pursuitmeter. The rested *Ss*, who had played with the pursuitmeter sufficiently previous to the experiment to know how it worked, were given as many 20-sec. trials on it, using the level of aspiration technique, as a four-min. interval permitted, and were rewarded for their performance no matter what it was. That is, they were told that they had surpassed their levels of aspiration, and that they were doing very well, much better than most people, better than they had done at card-sorting etc.

Three groups of 14 *Ss* each were run using these two tasks according to the procedure outlined in Table I. They will be referred to as the Neutral-Rest (N-R) group, the Failure-Rest

TABLE I

DESIGN OF THE EXPERIMENT

N-R (*Neutral-Rest*)		
Perform card-sorting to 45 sec. level. No failure.	Rest 4 min. Pursuitmeter success.	10 card-sorting trials Neutral; no scores requested or reported.
F-R (*Failure-Rest*)		
Perform card-sorting to 45 sec. level. Induced failure.	Rest 4 min. Pursuitmeter success.	10 card-sorting trials Neutral; no scores requested or reported.
F-NR (*Failure-No Rest*)		
Perform card-sorting to 45 sec. level. Induced failure.	No Rest	10 card-sorting trials Neutral; no scores requested or reported.

(F-R) group, and the Failure-No Rest (F-NR) group. The *Ss* in the two failure groups never knew they had reached the 45 sec. level. When their real score hit 45 sec. or lower they were stopped with the comment that it looked as if they would never reach this goal, that the *E* could not waste any more time on them, and that therefore they would go on to the next part of the experiment. In the Neutral-Rest condition the *Ss* did not experience much failure as they were allowed to reach their level of aspiration on every two out of three trials and to surpass it on the third. They also knew when they had reached the 45 sec. level and were stopped with the comment that they had done about as expected.

All *Ss* were told in advance what the experimental procedure would be, so that they could turn to the pursuitmeter task or the ten after-trials with a minimum of disruption and lost time. In the after-trials the levels of aspiration were no longer requested, no comments of any sort were made, and no scores were reported to the *Ss*.

All *Ss* were male undergraduates at Wesleyan University. The junior author, also an undergraduate, was the *E*. The senior author listened in on a number of *Ss* through a concealed microphone hookup. The rationale of this procedure and more details as to its *modus operandi* are given elsewhere (8).

RESULTS

Table II presents data which permit comparisons as to the initial equality of the three groups in learning ability. Since no failure had been experienced prior to the first trial the average number of seconds taken to sort the deck should be the same for all groups. Unfortunately this is not the case. There is one difference which is reliable—that between the Neutral-Rest and the Failure-No Rest group (Diff. = 13.14 sec., C.R. = 2.14, $P < .05$).[3] After the first trial the *Ss* in all groups were told that they had reached their level of

[3] *P* is the number of times in a hundred a difference that large would be obtained by chance. C.R. stands for critical ratio throughout.

TABLE II

AVERAGE SCORES MADE BY THREE GROUPS IN INITIAL TRAINING

Condition	N	Mean sorting time in sec.			Trials to reach 45 sec. criterion
		1st trial	2nd trial	Last trial	
N-R	14	79.64	67.29	43.64	10.21
σ mean		5.31	4.04	.40	.86
F-R	14	74.79	66.72	44.36	13.86
σ mean		4.75	3.53	.18	.80
F-NR	14	66.50	58.72	43.79	13.64
σ mean		3.07	2.58	.24	.93

aspiration. This means that averages on the second trial are also comparable. The comparison between the same two groups shows that the difference is now reduced to 8.57 sec. and is no longer reliable (C.R. = 1.79, $P < .10$). Nevertheless, there is good reason to believe that as a group the Failure-No Rest *Ss* were better card-sorters than those in the other two groups. Evidence in support of this point is to be found in Fig. 1 which shows the average number of

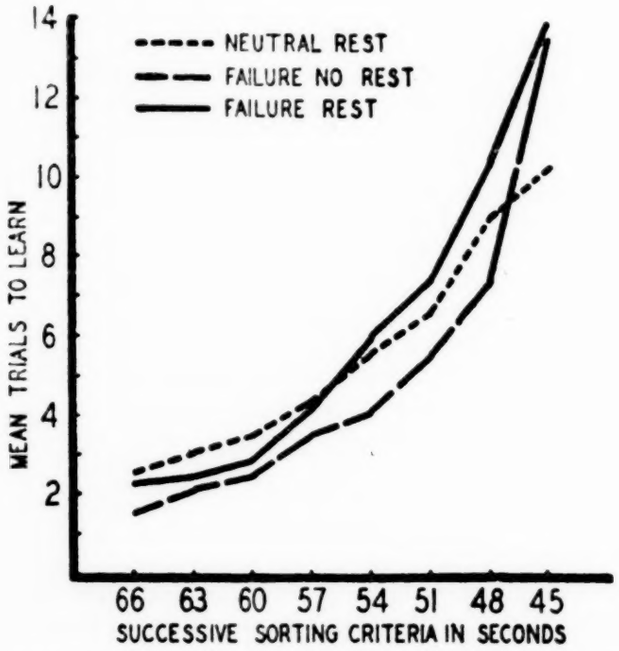

FIG. 1. Mean trails to reach successive sorting criteria in seconds for two failure groups and one neutral group.

trials taken by each group to reach successive sorting time criteria in seconds. The curve for the F-NR condition is consistently superior to that for the F-R condition up to the final 45 sec. criterion although none of the differences at any criterion is entirely reliable. It is superior even to that for the *Ss* who were not failing up to the last criterion. At this point the three groups are in the positions desired,

however. All have reached the same level of learning (see column 4, Table II) but the two groups which have experienced failure have taken about the same number of trials more to reach this level than the group which has not experienced failure (see column 5, Table II). The differences are reliable—that between N-R and F-R being 3.65 trials (C.R. = 3.09, $P < .01$), and that between N-R and F-NR being 3.43 trials (C.R. = 2.70 $P < .02$). The explanation for the fact that it took the superior F-NR group just as long to reach the 45 sec. level as it did the F-R group lies in the greater hostility which the *E* showed toward this group. Since his aim was to equate the two failure groups as to number of trials taken to reach the criterion, he had to adjust his level of hostility upwards with the better group to make the two come out the same. Although this made the two groups not strictly comparable in some respects, it had the advantage of providing two intensity levels of frustration, reactions to which are compared in another report (8).

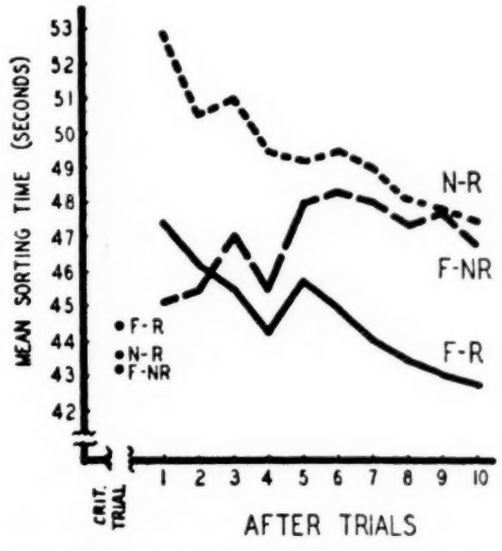

FIG. 2. Mean score (in sec.) per trial for the 10 trials after the original learning criterion had been reached. Two groups rested and one not rested.

It should be noted that at the 45 sec. criterion the two failure groups have had about three and a half more practice trials than the neutral group, but the level of mastery as shown by performance is the same. The hypothesis is that the failure groups have actually learned more from their added practice and that they are performing below their true level because of the conflicts introduced into their performing by inducing anticipations of failure.If this is so,removal of the conflicts by rest with success should permit the failure *Ss* to demonstrate their real superiority which has up to now been masked by inhibition.

The results of the performance of the three groups in the 10 after-trials are summarized in Fig. 2. Table III supplements Fig. 2 by

TABLE III

COMPARISON OF MEAN INITIAL AND FINAL SORTING TIMES IN SEC. FOR THREE GROUPS OF 14 *Ss* EACH IN THE PERIOD FOLLOWING THEIR DIFFERENT TREATMENTS

	First trial	Tenth trial
Neutral-Rest	52.79	47.36
σ mean	1.51	.90
Failure-Rest	47.36	42.57
σ mean	1.16	.67
Difference	5.43	4.79
σ diff.	1.91	1.12
Critical ratio	2.84	4.28
Failure-Rest	47.36	42.57
σ mean	1.16	.67
Failure-No Rest	45.00	46.79
σ mean	.50	1.14
Difference	2.36	4.22
σ diff.	1.27	1.32
Critical ratio	1.86	3.20
Neutral-Rest	52.79	47.36
σ mean	1.51	.90
Failure-No Rest	45.00	46.79
σ mean	.50	1.14
Difference	7.79	.57
σ diff.	1.59	1.45
Critical ratio	4.90	.39

supplying the mean scores of the three groups on the first and last trials, the differences among them, and the tests of their significance. It is first apparent that all three groups performed less well on the first after-trial than they had on the preceding criterial trial. A comparison with the scores in Table II shows that the N-R group lost on the average 9.15 sec. (C.R. = 5.87), the F-R group 3.00 sec. (C.R. = 2.56, $P < .02$) and the F-NR group 1.21 sec. (C.R. = 2.09, $P < .05$).

If the two groups which were rested are compared, the failure group lost very much less. It performed reliably better on the first after-trial than the neutral, rested group. The difference of 5.43 sec. is about equal to the amount gained in three and a half trials of practice at this level of learning, a fact which can be gathered from inspection of Fig. 1. Apparently removal of conflict by success enabled this group to display a superiority which had formerly been disguised. Comparison of the two failure groups shows that the introduction of a rest did not improve sorting time as compared with a no-rest group. If anything it was detrimental to performance. The critical ratio (1.86, $P < .10$) does not meet the highest criterion of significance, and should probably be further discounted in view of the probable superior sorting ability of the F-NR group. A safer conclusion seems to be that these two groups performed about the same on the first after-trial.

The next point of interest is the general slope of the curves. The two rested groups improve steadily and at about the same rate, so

that the two curves are practically parallel, confirming the impression that the failure group is now displaying an advantage gained by extra practice in initial learning. The improvement from the 1st to 10th trials is reliable in both cases, being 5.43 sec. (C.R. = 3.16) in the N-R group and 4.79 (C.R. = 4.20) in the F-R group. The curve for the non-rested failure group displays entirely different characteristics. The *S*s get steadily worse up to the sixth trial, when there has been a loss of 3.29 sec. from the first trial (C.R. = 3.70). Thereafter they show a slight improvement, which makes the difference between the first and tenth trial smaller (1.79 sec.) and less significant (C.R. = 1.77, $P < .10$).

On the tenth and last trial the two rested groups still differ significantly as they did on the first trial, but the F-NR group has changed its position. It now performs significantly less well than the failure group which was rested with success.

Discussion

An interpretation of the results of this experiment can perhaps best be made by paraphrasing the theory that Hull (**3**) has introduced. His basic assumptions are that a certain amount of excitatory and inhibitory potential accumulate during any learning, and that actual performance or habit strength is a function of excitation minus inhibition (cf. **3**, pp. 300, 383). Normally, in an experiment, inhibition is not controlled. Its existence is inferred and usually attributed to conflicts developed in the course of learning. This was the approach used by the senior author in his previous studies (**6, 7**). In the present experiment, conflicts which were known to interfere with performance (**10**) were introduced to create a large amount of inhibition experimentally. The *S* was made to fail. As a result he began to anticipate failure and responses like "Oh damn, there I go again" began to appear and to interfere with his performance. Under these conditions *S*s took longer to reach a specified level of habit strength than those not made to fail. During this added practice period they presumably accumulated sufficient excitatory potential to outweigh the greater inhibition introduced into their learning by the *E*. To demonstrate the presumed existence of greater excitatory potential in the F-R than in the N-R group, the inhibition in both groups was reduced by success in a rest and the "true" superiority of the failure group showed up. As for the failure group, which did not have the inhibition removed by success, their conflicts continued through the after-trials, eventually causing a deterioration in score, possibly because the conflicts were more intense now that the possibility of success was removed, or possibly because drive which operates in conjunction with habit strength (**3**) was lowered.

While the superior performance of the F-R group over the N-R group shows that rest may operate differentially on two groups which have reached the same level of learning, it does not constitute an instance of reminiscence as usually defined (1), since it is not superior to the performance of a similar group after no rest. The explanation of this is not hard to find, however. In both failure groups excitatory and inhibitory potential are larger than in the N-R group, though their difference, i.e., their actual performance, is the same. In order to reduce inhibition, success on a second motor task is introduced in one failure group and not in the other. It is reasonable to assume that, besides reducing inhibition, this task interfered with the first causing retroactive inhibition.[4] Such interference would not influence the comparison of the two rested groups since its detrimental effect would be equal in both. But it would tend to lower the score of a rested group as compared with one that had had no rest and had not performed the second task. Consequently, the F-R group should be worse than the F-NR group because of retroactive inhibition, and better because of removal of inhibition. Apparently the two effects cancelled each other approximately and the scores of the two groups on the first after trial are about the same. If the inhibition could have been removed without resort to a second task, the F-R group should be superior to the F-NR group and should show true reminiscence.[5]

So far it has been assumed that the theory which led to the design of the experiment is the best one to explain the results. Are there other interpretations which will explain them equally well? It is difficult to account for the results in any very meaningful way by the differential rate of forgetting theory. It would have to be assumed that failure somehow caused the *S* to fixate (i.e., acquire) more wrong responses than non-failure and the success during a rest somehow facilitated the foregetting of these wrong associations. If forgetting is largely the result of retroactive inhibition, success on the pursuit-meter would have to cause forgetting by competing with wrong responses in card sorting. Futher analysis may make this interpreta-

[4] Although Buxton and Henry (2) did not find evidence of retroaction in pursuit learning, the fact that there is considerable loss for both rested groups, whereas they found only gains shows that card sorting is not in the same class. It is apparently not so entirely a motor task and is consequently subject to retroaction as are most other activities (cf. 9).

[5] Confirmation of this hypothesis is to be found in Sears curves (10, p. 575). His failure *Ss* did progressively less well on a given day, but between days there was marked reminiscence; that is, the first three trails on the next day are markedly superior to the last three trials on the preceding day. Apparently the anticipation of failure which interfered with performance disappeared in the 24–48 hour period between practice periods without any special success experience. These data are not conclusive because there was no control group, the retention of which could be compared with the group which had failed, but they point in the same direction as the results obtained here.

tion more plausible, but the chief difference between it and the one adopted here can be clearly stated now: in the differential rate theory interfering responses are *learned* (acquired) during the experiment proper and *unlearned* (forgotten) during rest. In the removal of inhibition theory the interfering responses are *evoked* (learned at some other time) through generalization (cf. 7) or a conflicting motive (in this instance, fear of failure) in original learning and are *not evoked* in relearning because of some change (i.e., differentiation or change of motivation) in the state of the *S* during rest.

Two other possible interpretations suggest themselves. In the first place, suppose that there really was no difference in true learning (or excitatory potential) between the neutral and failure groups, that their scores on the criterial trial represented the true state of affairs. The question then becomes: is there any factor in their identical treatment during rest which might act differentially on the two groups producing a very different score on the first after-trial? It is possible that the success led to an increased effort on the part of the *S*s who had been failing. They might have tried harder now that they knew they could succeed. That is, the difference might lie in a shift in drive rather than in a removal of inhibition. There are several difficulties with this explanation, however. For one thing, there should be less effect on the score on the first trial than later on. One would expect the increased motivation to show its effect slowly, so that the two curves would gradually diverge instead of running parallel as they do. For another, there was no particular reason to try to do better since the *S*s were no longer told their scores and did not know whether they were improving or not.[6]

There is also the possibility that it was the removal of the requirement that the *S*s state a level of aspiration which made the difference in the score. That is, the failure *S*s may have been relieved now that one possible source of failure had been removed, whereas the neutral

[6] Actually, there may be very little difference between this interpretation of the facts and the one based on removal of failure-induced inhibition. A more thorough analysis would suggest that the failure *S*s find the card-sorting less and less rewarding during original learning, that they continue gaining but do not show all their gain because a new motive—to avoid failure—tends to evoke responses (anticipations of failure) which interfere with rapid performance of the task. During rest the motive to avoid failure is weakened by giving the *S*s more success and self-confidence. Consequently the interfering responses which this motive has evoked in original learning no longer appear and there is a decided improvement in performance. This interpretation joins the motivational shift theory with the removal of inhibition theory by explaining the latter in terms of the former. Another way to saying it is that the interfering responses inhibit learning initially but that these responses are in turn dependent on the strength of a conflicting motive. The interfering responses are the *mechanism* by which inhibition is produced, the motive is the *cause* of the mechanism. However, this explanation still assumes that the failure group had learned more in the original practice than the neutral group. No shift in motivation theory can readily account for the large differential in score appearing on the first after-trial without assuming some latent differential in original learning caused by some kind of inhibition in the failure group.

Ss would be unaffected. The difficulty with this theory is that removing the requirement for the failure *Ss* who had no rest certainly did not lead to the same kind of after-trial performance as for the rested *Ss*. Something which happened in the rest must have been the crucial factor.

The alternatives to the removal of inhibition explanation are not only dubious in themselves. There is strong positive evidence from the remarks of the *Ss* to indicate that the success did remove conflict. The record of remarks made by the *Ss*, often during sorting, showed that the faliure *Ss* swore a good deal during the initial training. They also fumbled and misplaced the cards. All such signs of conflict were greatly reduced in the after-trials. It seems reasonable to assume that they could sort faster because such interfering emotional responses were removed. That the success in the rest actually served this purpose is attested by such remarks as: "Oh boy, maybe that will help my first failings" and "I needed that little stimulation."

Further support is to be gained for the removal-of-inhibition theory from the fact that the difference between the two rested groups on the first after-trial is about equal to the gain in score which a group would normally have made which had practiced three and a half extra trials. Since this is the amount of extra practice which the failure group had had, it seems reasonable to assign the advantage it shows to this factor. The situation is somewhat comparable to latent learning where an organism acquires something which does not appear until the conditions have changed so as to allow it to do so.

Summary

Three groups of 14 male undergraduates were asked to perform a card-sorting task until they could sort through the standard pack in 45 sec. All were asked to state a level of aspiration before each trial, and a score (purporting to be correct) was reported to them all after each trial. To one group the *E* reported scores such that each *S* reached his level of aspiration about two out of every three trials and surpassed it on the third; to the two other groups he reported scores such that each *S* failed to reach his level of aspiration on about two out of three trials and just reached it on the third. He also swore at the *Ss* in these groups for their failure and stopped them when their real score reached 45 sec. with the remark that they would never make it. The neutral group and one failure group were then given a short rest during which they practiced with success on a pursuitmeter. The other failure group proceeded immediately to 10 further trials required of all groups in which no levels of aspiration were requested, no remarks made, and no scores given out. The following results were obtained.

1. The two failure groups required a reliably greater number of trials to reach the 45 sec. criterion than did the neutral group.

2. Although all three groups sorted equally fast on the criterial trial, marked differences appeared on the first after-trial. The rested failure group was reliably superior to the rested neutral group and equal or slightly inferior to the non-rested failure group.

3. The performance curves for the 10 after-trials shows slight improvement at the same rate for the rested failure and neutral groups, so that the two curves remained approximately parallel, with the former maintaining its initial superiority. The non-rested failure curve grew reliably worse in the after-trials.

4. The most satisfactory explanation of these results was found to be that failure produced conflicts in responding which lowered performance below the true level of learning. Removal of these conflicts by rest permitted the superiority of the failure group, which was gained in the extra original learning trials, to show itself in comparison with a similarly rested group which had had no performance conflicts induced. True reminiscence was not obtained, probably because the practice during rest necessary to give the feeling of success produced retroactive inhibition, as well as the desired removal of conflict.

The experiment supports the senior author's hypothesis that reminiscence is due to the removal of inhibition by providing an instance in which a superiority in performance after no practice is to be attributed to inhibition experimentally introduced and withdrawn.

(Manuscript received June 10, 1946)

References

1. Buxton, C. E. The status of research in reminiscence. *Psych. Bull.*, 1943, **40**, 313–340.
2. Buxton, C. E., & Henry C. E., Retroaction and gains in motor learning. I. Similarity of interpolated task as a factor in gains. *J. exp. Psychol.*, 1939, **25**, 1–17.
3. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
4. McClelland, D. C. Reminiscence as a performance phenomenon. Unpublished Departmental Seminar Memorandum, Yale University, 1941.
5. McClelland, D. C. Studies in serial verbal discrimination learning. I. Reminiscence with two speeds of pair presentation. *J. exp. Psychol.*, 1942, **31**, 44–56.
6. McClelland, D. C. Studies in serial verbal discrimination learning. II. Retention of responses to right and wrong words in a transfer situation. *J. exp. Psychol.*, 1942, **31**, 149–162.
7. McClelland, D. C. Studies in serial verbal discrimination learning. III. The influence of difficulty on reminiscence in response to right and wrong words. *J. exp. Psychol.*, 1943, **32**, 235–246.
8. McClelland, D. C., & Apicella, F. S. A functional classification of verbal reactions to experimentally induced failure. *J. abnorm. soc. Psychol.*, 1945, **40**, 376–390.
9. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, 1942.
10. Sears, R. R. Initiation of the repression sequence by experienced failure. *J. exp. Psychol.*, 1937, **20**, 570–580.

THE ILLUSORY PERCEPTION OF MOVEMENT CAUSED BY ANGULAR ACCELERATION AND BY CENTRIFUGAL FORCE DURING FLIGHT. I. METHODOLOGY AND PRELIMINARY RESULTS

BY CAPT. ASHTON GRAYBIEL, (MC) USN, LT. COMDR. BRANT CLARK, H(S) USNR AND ENS. KENNETH MACCORQUODALE, H(S) USNR

School of Aviation Medicine Naval Air Training Bases Pensacola, Florida *

Studies in this laboratory have demonstrated three forms of illusory perception of movement which might be experienced by a pilot flying at night: namely, the auto-kinetic illusion, the oculo-gyral illusion and the 'oculo-gravic illusion.'

The autokinetic illusion is well known and refers to the apparent wandering of an object in the field of vision; the immediate cause is unknown, but an important factor is a lack of adequate visual reference cues. A previous study (1) has shown that this illusion may be observed during formation flights at night if one or even more lights on the leading plane are fixated.

The oculo-gyral illusion (2, 3) refers to a form of apparent movement resulting from stimulation of the sensory receptors in the semicircular canals following angular acceleration. This familiar illusion, sometimes referred to as 'visual vertigo,' may be observed in daylight but is more prominent at night. Threshold studies (4) soon to be reported from this laboratory show that angular accelerations well below those frequently encountered during flight maneuvers are sufficient to provoke this illusory phenomenon.

The oculo-gravic illusion (5) refers to the apparent displacement of an object in space which may be observed when the sensory receptors in the otolith organs are stimulated by an accelerative force which forms a resultant vector with the force of gravity. Experiments on the human centrifuge have shown that if man is subjected to accelerative force, he will orient himself to the resultant of this force and the force of gravity, and that if visual cues are limited he will localize an object in space in accordance with this new orientation.

These three illusory phenomena have been studied separately in the laboratory but only one of them, autokinesis, has heretofore been

* The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the view of the Navy Department or the Naval Service at large.

studied during flight (1). It remained, therefore, to study the illusory effects during various maneuvers in which both angular acceleration and *g* forces are acting on the flyer. Inasmuch as both angular acceleration and *g* may be acting simultaneously, and, in addition, autokinesis may be present as well, the apparent movement observed may represent a global effect. The purpose of this report is to describe a method which is useful in studying these illusory movements during flight. The results obtained in preliminary experiments are briefly described.

Apparatus and Procedure

The apparatus used to investigate the effects of *g* and angular acceleration on visual perception during flight was mounted in an SNJ-6 aircraft. This is a standard two-place Navy training plane. The pilot sat in the front and the observer in the rear cockpit. Two special pieces of equipment were mounted in the rear cockpit where they could be used and controlled by the observer.

Fig. 1. Apparatus and subject in place in rear cockpit. Official U. S. Navy Photograph

A standard General Electric Wire Recorder was wired into the radio equipment in such a fashion that, by means of a selector switch, it could be made to record either continuously or when a microphone switch was depressed. In either case all of the pilot's or observer's reports would be automatically recorded. When the recorder ran continuously, the wire provided a self-contained time line for the recorded events.

A collimated star was also mounted in the rear cockpit (Fig. 1). The star was contained in a cylindrical, light-proof housing and was viewed through a four-in. collimating optical system. The star consisted of a luminous point with six equally spaced radiating lines. The device was set on a heavy rubber base which damped out the plane's vibration and gave a clear test object. The base was mounted in turn on a movable carriage which rode on a semicircular gunnery track located in the cockpit directly in front of the observer. Thus, the star could be easily fixed at any position along a 180° arc in front of the *S* from his right to his left. Throughout the observations reported here, the fixation object remained stationary directly in front of the observer.

The circular star housing was fitted with an opaque sleeve to which was attached a standard heavy goggle mounting. To make observations the observer put on the goggle and thus could observe the star in darkness. The black-out was made complete by means of a heavy sateen cloth which covered the instrument and the observer's head throughout the observations. A bite-board attached to the star housing was used to fix the *S*'s head to eliminate any relative motion between the *S* and the test object, but due to the vibration of the plane, it was held in the mouth without tightly clamping the teeth. This arrangement allowed the *S* to sit in the rear cockpit and observe an isolated fixed test object in the dark at an optically infinite distance, while the pilot maneuvered the plane in daylight.

The observer, using a throat microphone, gave a running account of the apparent motion of the star during each maneuver. In each case he reported three things: (a) motion, (b) displacement, and (c) rotation of the test object. The illusory movements of the star made it necessary to make a sharp distinction between rate of motion of the target on one hand and the amount of displacement on the other (3). Motion here refers to that perception which defines a body as not being at rest, while displacement refers to a change in location of the object in relation to the observer. This distinction was necessary, since apparent motion was observed with and without associated displacement of the target and, once displaced, the target might remain stationary in its new location. Rotation refers to a turning about the center point in the plane of the radiating lines. Apparent motions and displacements of small magnitude were disregarded. Direction of motion was indicated by reference to the numbers on the face of a clock, e.g., if the object appeared to move to the left, the *S* reported 'nine.' The amount of apparent displacement and of apparent rotation were roughly indicated in degrees from an egocentrically located 'true' position of the star. Certain prearranged verbal signals were used to indicate the cessation of movement and displacement. The presence or absence of these three types of change was reported periodically. This served to center the attention of the observer on all of the types of motion which were observed, since preoccupation with one type of change or fluctuation of attention tended to cause the observer to disregard the other types of apparent motion.

The pilot indicated critical points in each maneuver by means of a time signal on the wire using a phonetic alphabet. For example, each trial was begun by the pilot, and the beginning point was indicated by his saying ABLE into the microphone. Subsequent gross changes in attitude were similarly identified on the wire by the pilot's saying BAKER, CHARLIE, etc.

Each experimental series was planned in advance to consist of 10 maneuvers in addition to two control periods during which time straight and level flight was maintained. The order of the trials was prearranged in random order by the pilot. Hence, the order of presentation was unknown to the observer and was different in each series.

Each maneuver began from straight and level flight at 130 mph and terminated after one min. of straight and level flying following recovery from the maneuver. Precise descriptions of the stimulus-response relationships are not possible, since measuring devices were not available to indicate the exact accelerative forces involved. Therefore, the stimulus situation is described in terms of the maneuver itself together with a general description of the accelerative forces. Rest periods were taken after each third or fourth trial. A preliminary investigation has been made of the following maneuvers:

(1) *Climb and glide.*—The pilot gradually raised or lowered the nose until he attained a rate of climb or glide of 500 ft. per min. This was held until the plane had gained or lost 500 ft., at which time the pilot gradually leveled off. In each case there was a loss in air speed of 20 mph. In the climb there was a temporary slight increase in *g* at the beginning of the climb and decrease in *g* during recovery. The reverse was true of the glides. There was a small angular acceleration about the lateral axis of the plane in both the climbs and glides as the plane assumed the new attitude.

(2) *Turns.*—The pilot gradually lowered the wing to attain an 18° bank and maintain a rate of turn of 3° per sec.; upon completion of a 180° turn he gradually returned to straight and level flight. As the plane assumed the turning attitude there was a small angular acceleration in the direction of turn and an 19° tilt about the longitudinal axis. There was a slight increase in *g* due to centrifugal force (theoretically 1.05 *g*). The change in air speed was negligible.

(3) *Climbing turn (chandelle).*—The pilot dropped the nose to gain a speed of 175 mph and then began a sharp climbing turn. When he had completed a 180° turn, he gradually resumed straight and level flight. There was a small angular acceleration around the lateral axis as the

pilot dropped the nose of the plane. At the beginning of the climbing turn there was a sharp increase in *g*, angular acceleration in the direction of the turn, and angular acceleration in the longitudinal axis due to the bank.

(4) *Diving turns.*—The pilot lowered the nose 60° below the horizon and turned sharply. When the plane had lost approximately 1200 feet, he gradually recovered to straight and level flight. There was an initial small angular acceleration about the lateral axis of the plane at the beginning of the dive. As the plane assumed the turning attitude there was a small angular acceleration in the direction of turn, and a sharp increase in *g*. There was also angular acceleration about the longitudinal axis as the plane banked. The increase in air speed was approximately 60 mph.

(5) *45° Rolls.*—The pilot raised the wing and rolled 45°. He then rolled back through center and to 45° in the opposite direction. He then recovered to straight and level flight. The angular accelerations were about the longitudinal axis of the plane. There were no appreciable changes in *g* or in air speed.

(6) *Controls.*—Straight and level flight was maintained throughout these trials. The pilot indicated prearranged time intervals over the microphone. These time intervals approximated the times used to indicate critical events in the other maneuvers.

A typical experimental run involved approximately one hour of flying time. The pilot chose an altitude of between 4000 to 8000 ft., depending on atmospheric conditions. The observer would then fixate the star, adjust the blackout-cloth and, after labeling the wire and setting the wire recorder to run continuously, indicate that he was ready for the first trial. Following the pilot's signal, ABLE, the observer would report the apparent behavior of the star until the end of the trial. The observer then made any additional comments that were necessary to clarify the record. In some trials no motion, displacement, or rotation were observed. Nevertheless, during the trials the observer would periodically indicate that there was no motion. On the other hand, during some maneuvers like a chandelle or diving turn events happened so rapidly that it would be necessary to report every two or three sec. Recording the apparent motion and displacement of the star during such maneuvers was a highly complex task requiring a considerable amount of practice. Therefore, it was impossible to use inexperienced observers.

The three *S*s who made the observations were the authors. All three had had considerable laboratory experience in observing and reporting the effects of angular acceleration and *g* on a fixed target in the dark (3). However, these were their first observations of these effects in flight.

All data from the wire recorder were transcribed into a written protocol in the laboratory. All time relationships were measured from the pilot's signal, ABLE. The wire recording was played back one or twice to record the statements and the approximate times. In addition, as a final check two persons listened together and made independent time checks to the nearest second. If the two failed to agree, the trial was rerun and additional checks were made until they were able to agree. The analysis of the data was then made from these protocols.

Results and Discussion

Preliminary data have been gathered on the standard maneuvers listed above and are described here as samples of the types of illusory perception involved. Data for maneuvers to the right and left are combined. The nystagmus resulting from rotation to the right or left will be expected to result in apparent motion in the direction of turn (3). Therefore, the analysis of the apparent motion has been in terms of its correspondence to the direction of turn.

The control runs, during which the pilot maintained the plane in straight and level flight, were largely uneventful. During the 19 trials totaling 49 min. of observation, two *S*s reported six brief movements of a total duration of 24 sec. In terms of percentage, this was 0.85 percent of the total observation time. There was no tendency for any of the reports of movement given to stand in relation to the

pilot's signals; nor were the movements preponderant in any direction. Both the characteristics of these movements and the circumstances under which they were observed strongly suggested that they were autokinetic in nature. Such observations of apparent movement as were made were equally divided between the two observers. There were no reports of displacement or torsion of the test object. It is apparent that the star itself, suggestion, or normal operation of the plane in straight and level flight were not adequate stimuli to produce the effects observed when actual maneuvers were performed.

Movement was reported in 18 out of the 20 standard rate, 180° turns. The onset of apparent movement in nine of the 18 instances was observed before the plane had completely assumed the turning attitude, and in the remaining nine instances at varying times before the turn was completed. The direction of apparent movement varied: in 8 instances it was predominantly upward; in 4 downward; in 2 it was opposite to the direction of turn; and in 4 it was in the same direction as the turn. On two additional trials a movement was observed subsequent to the original motion and in the opposite direction. Displacement upward was reported in seven of the trials. Movement was seen during 11 out of 20 recoveries to straight and level flight. The star appeared to move downward in five trials and opposite to the direction of turn in six trials. The displacements during recovery tended to return the star to its original position.

During climbs and glides motion was observed in the vertical meridian only. At the beginning of both climbs and glides, there was a tendency to see motion downward more frequently than upward. However, on recovery from the climbs there was a marked tendency to see motion downward and on glides to see motion upward. Motion was observed in 14 out of 22 climbs and glides and in 20 out of 21 recoveries from climbs and glides.

The climbing and diving turns gave rise to the most striking illusions of movement and displacement. In these maneuvers the *g* forces were substantial. Illusory motion and displacement occurred in all of the trials. Events occurred so rapidly that considerable difficulty was encountered in making a complete verbal report. In fact it was quite impossible on the first two or three attempts. During the initial dive in the chandelle and the first part of the dive in the diving turn, the light usually appeared to move downward. As the climbing or diving turn became established and the *g* force increased, the star appeared to move and at the same time be sharply displaced upward. The motion was rapid and the maximum displacement was usually between 20° and 30°. One apparent displacement of 60° was recorded during a climbing turn. During the chandelle, one of the observers also reported a slight veering of the

upward movement toward the direction of turn. This occurred in five out of eight trials. Recovery from both the climbing and diving turns resulted in simultaneous motion and displacement to the normal position of the star. This occurred in all of the trials.

The 'quarter roll' is not a coordinated maneuver, so the *S* was exposed to displacement in the seat throughout the trial. This confused the report in many cases, since it interposed marked extraneous cues. These reports were so difficult to make that no approximations, however rough, of time relations can be made at this point. Movement of some sort was reported in fourteen out of nineteen trials. These involved: clockwise and counter-clockwise rotation of the star and movement in the direction of bank. No displacement was reported.

These preliminary results show that the method developed can be used to study the effects of the stimulation of the semicircular canals and otolith organs on visual perception during flight. Apparent motion which can be attributed to the nystagmus associated with rotation was observed in the standard turns and to a lesser degree in the climbing and diving turns. Motion and displacement which may be attributed to stimulation of the otolith organs was observed particularly during the climbing and diving turns. Since a precise description of the stimulus situation is not possible and since only preliminary results are available, it would be precipitate to attempt a more elaborate description of the stimulus-response relationships at this point.

Marked variability in response occurred not only between *S*s but also on different trials for the same *S*. At this time it is not clear how much of this is due to the operation of the aircraft and how much is a result of individual variability of response. Autokinesis cannot be completely ruled out as a contributing factor in these results. However, since the illusory perceptions were related to the angular acceleration and *g* forces, which were known to be operating, autokinesis was not an important contributing factor. This is further borne out by the small amount of apparent motion (autokinesis) in the control runs.

This method of studying visual orientation in the air has several unique advantages. It makes it possible to re-examine phenomena observed in the laboratory in a practical situation. It provides a situation wherein the *S* is free in space to move in any direction. A wide range of linear and angular accelerations and *g* forces are encountered. The practical significance of the results as they concern aviation is assured by collecting the data in actual flight.

Extensive controls of the experimental situation are required if reliable data are to be obtained. Trained subjects only must be

used. This is particularly true in maneuvers such as a chandelle, where strong *g* forces make reporting complicated and where preoccuptation with other, if not unrelated, bodily sensations may invalidate the results. Plane vibration contributes an irrelevant motion of the visual stimulus, but was found to be a minor problem when the star was suitably mounted. The observer's head should remain motionless, but the usual cumbersome apparatus for immobilizing the head was impractical; the bite-board was a satisfacotry solution. It was necessary to seek smooth air, but the selection of the proper altitude solved this problem. The rough air from the plane's own slip steam in rapid turns caused only a momentary interruption. It was necessary to control certain extraneous cues, such as cues to *g* derived by lifting the feet or cues to the direction of turn from the sun through the canopy falling on the shoulder. It was impractical to attempt to control the auditory cues resulting from changes in power setting during the maneuvers. It was discovered, however, that, in general, engine noise and plane vibration for different maneuvers did not differ characteristically, and therefore did not provide adequate cues for accurately differentiating between them. The value of control runs could be increased by adding power at appropriate intervals.

The stimulus situation, in terms of the maneuvers performed and the resultant changes in direction of force on the observer, requires careful planning and some knowledge of aerodynamics. In maintaining a turn of any given degree, for example, if a constant power setting is used the angle of attack of the wing must be increased, with a resultant change of direction of the resultant force of *g* on the otolith organ. If, on the other hand, the angle of attack is kept constant it is necessary to increase the speed.

It is proposed to study the following types of problems with this experimental method: the determination of thresholds for the influence of angular acceleration and *g* on visual perception, and thresholds for the perception of rotation and of *g per se*.

Summary

A method was evolved to observe and report the effects of angular acceleration and variations in *g* on visual perception during flight. The visual stimulus was a collimated 'star' installed in the rear cockpit of a standard Navy training plane. All observations were made in complete darkness. Both the pilot's and observer's verbal reports were dictated into an airborne wire recorder which also provided a time line. These recordings were transcribed in the laboratory and all analyses were made from these transcriptions.

Preliminary experiments were carried out on three Ss using five basic maneuvers in addition to control periods of straight and level flight. For the first time it has been shown that angular acceleration and *g* during flight induce illusory perceptions of motion and displacement of an objectively motionless object. The magnitude of the illusory effects may be great, the fixation object appearing to be displaced as much as 60° from its actual position. The significance of this finding for aviation is evident. The method is adaptable to a number of types of studies concerned with the influence of angular acceleration and *g* on visual perception.

We gratefully acknowledge the assistance of Lt. Joseph Achin, USNR, who served as the pilot.

(Manuscript received May 20, 1946)

Bibliography

1. Graybiel, A., & Clark, B. The autokinetic illusion and its significance in night flying. *J. Aviat. Med.*, 1945, **16**, 111–151.
2. Graybiel, A., & Hupp, D. I. The oculo-gyral illusion: a form of apparent motion which may be observed following stimulation of the semicircular canals. *J. Aviat. Med.*, 1946, **17**, 3–27.
3. Graybiel, A., Clark, B., MacCorquodale, K., & Hupp, D. I. The role of vestibular nystagmus in the visual perception of a moving target in the dark. *Amer. J. Psychol.*, 1946, **59**, 259–266.
4. Graybiel, A., Kerr, W. A., & Hupp, D. I. Thresholds of stimulation of the horizontal semicircular canals in man. (In preparation)
5. Graybiel, A., Hupp, D. I. & Patterson, J. L., Jr. The law of the otolith organ. *Fed. Proc.*, 1946, **5**, 35.

THE RELATIONSHIP OF EXPOSURE TIME AND ACCURACY IN A PERCEPTUAL TASK

BY B. R. PHILIP

Fordham University

Introduction

Thurstone (**6**), on purely theoretical grounds, determined that the relationships between ability, motivation and speed could all be expressed by an integration of the normal probability curve. This deduction is partly at variance with the experimental findings as summarised by Garrett (**1**), who cited detailed data, in addition to his own, to show that in both motor and perceptual tasks there occurred an optimal time interval, generally near the center of the time series, at which accuracy was highest. According to his conclusions there is no simple relationship between speed and accuracy. Philip (**3**), on the other hand, found that this relationship might be expressed by an integration of the normal curve. His findings, while restricted to a motor task, are in accordance with the theoretical deductions of Thurstone, which were published later. Where perceptual tasks are concerned no experimental confirmation of the Thurstonian deduction has, to the present author's knowledge, since appeared, although the data of several papers, e.g., Hunter and Sigler (**2**) and Young (**7**), when reworked in the light of this problem, seem to confirm the theoretical deduction.

The present paper attempts to re-examine the problem of the relationship between exposure time and accuracy in a perceptual task. Specifically, it attempts to answer the question: does accuracy, in a perceptual task, diminish according to some continuous function with increasing speed of performance of that task, or is there an optimal time interval, somewhere between the highest and the lowest speeds, at which accuracy is at a maximum?

An experimental investigation of this problem presents some difficulties. Since small differences in the performance at various speeds are to be expected, for reliable results the perceptual material must be presented a great number of times. Suitable material must be carefully selected so as to be largely independent of prior experience and schooling. Scoring must be objective and practice effects in learning must be relatively slight. For purposes of this investigation material previously used by the author for the discrimination of color

mass (4, 5) seems to be suitable, since it fulfils all these conditions and in addition permits judgments to be made rapidly.

Two experiments, involving similar material, different *Ss* and different psychophysical methods, were carried out and will be individually described.

First Experiment

The material consisted of sets of 20 cards, $3\frac{7}{8}$ by $2\frac{3}{8}$ in., on each of which quarter-in. dots were printed in regular rows, using four different colors, blue, green, red and purple. One card had 18 blue dots and 6 each of the other colors. The color scheme of the other cards in which blue predominated was as follows: 16 blue dots, and 7, 7 and 6 of the others; 14 blue dots, and 7, 7 and 8 of the others; 12 blue dots, and 8, 8 and 8 of the others; 11 blue dots and 8, 8 and 9 of the others. This scheme was followed for each of the four colors. By this method five levels of difficulty of discrimination were obtained, corresponding to the various amounts of the predominating color (4).

The cards were presented singly in a modified Dodge tachistoscope, with a Marietta timing device set for a definite time interval. The instructions were simply to call out the predominating color—a task which may be readily and quickly done, as the *S* has an immediate perception of a color mass of a definite hue. Judgments of 100 cards constituted a session, during which each set of 20 cards appeared five times in randomized order. The order of presentation was previously determined, and the cards were carefully stacked according to this order which was listed on score sheets, upon which the *S*'s responses were recorded.

Five time intervals of exposure of the stimulus cards were used: .133, .200, .299, .447 and .668 sec.; but the exposure time was kept constant throughout a session. The order of selection of the exposure times was randomized on the basis of Tippett's tables, and each exposure interval occurred five times, so that there were 25 sessions in all. Thus, for each *S* there were presented 500 stimuli at each exposure time, making a total of 2500 judgments per *S*.

Each session, during which 100 cards were judged, lasted approximately 20 min., and the experiment was carried on over a period of some seven weeks, at the rate of about four sessions a week. Prior to each session, there was a warming-up period during which 20 cards were judged, although the judgments were not recorded. Before the experiment proper, two sessions were set aside for practice. Three graduate students, who were competent psychological observers, took part in this experiment.

Results

The numbers of errors made at each time interval are shown in Fig. 1; here accuracy diminishes, but not linearly, with exposure time. Since there were five sessions devoted to each exposure time, the practice effect may be noted by the variation in the number of errors made at each of these sessions. The total numbers of errors given by all three *Ss* for the various time intervals are listed in Table I. There is about a five percent reduction in error from session to session.

The data, fractionated according to the number of errors made at each of the five levels of difficulty, are presented in Fig. 2. The easiest level (material containing 18 dots of the predominating color), and the most difficult level (material containing 11 dots of the predominating color), do not show any sizable difference at their respective level for the various time intervals of exposure. Obviously,

the material at these levels of difficulty was not suitable for the purpose of the experiment. However, the cards of the three middle levels of difficulty do show, in conformity with Thurstone's theory, a functional relationship of the ogive type between exposure time and accuracy.

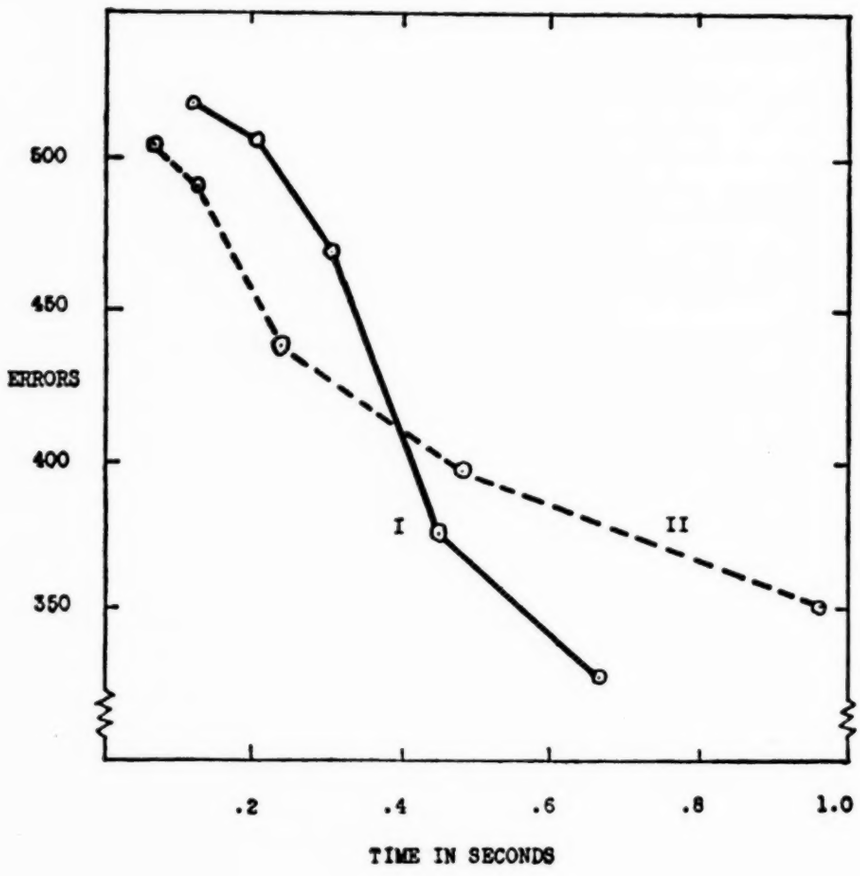

FIG. 1. Relationship between number of errors and exposure time. The solid line (I) refers to the first experiment, the dotted line (II), to the second experiment. The ordinates are erroneous responses for I and weighted errors for II; in the latter case the ordinate values given must be multiplied by ten.

TABLE I

NUMBER OF ERRORS AT DIFFERENT SPEEDS AND STAGES OF PRACTICE, TASK I

Sets	Exposure Time in Sec.					
	.133	.200	.297	.447	.668	Total
I	112	118	111	77	69	487
II	96	108	99	88	66	457
III	115	100	89	83	61	448
IV	104	84	95	71	70	424
V	91	98	78	61	60	388
Total	518	508	472	380	326	2204

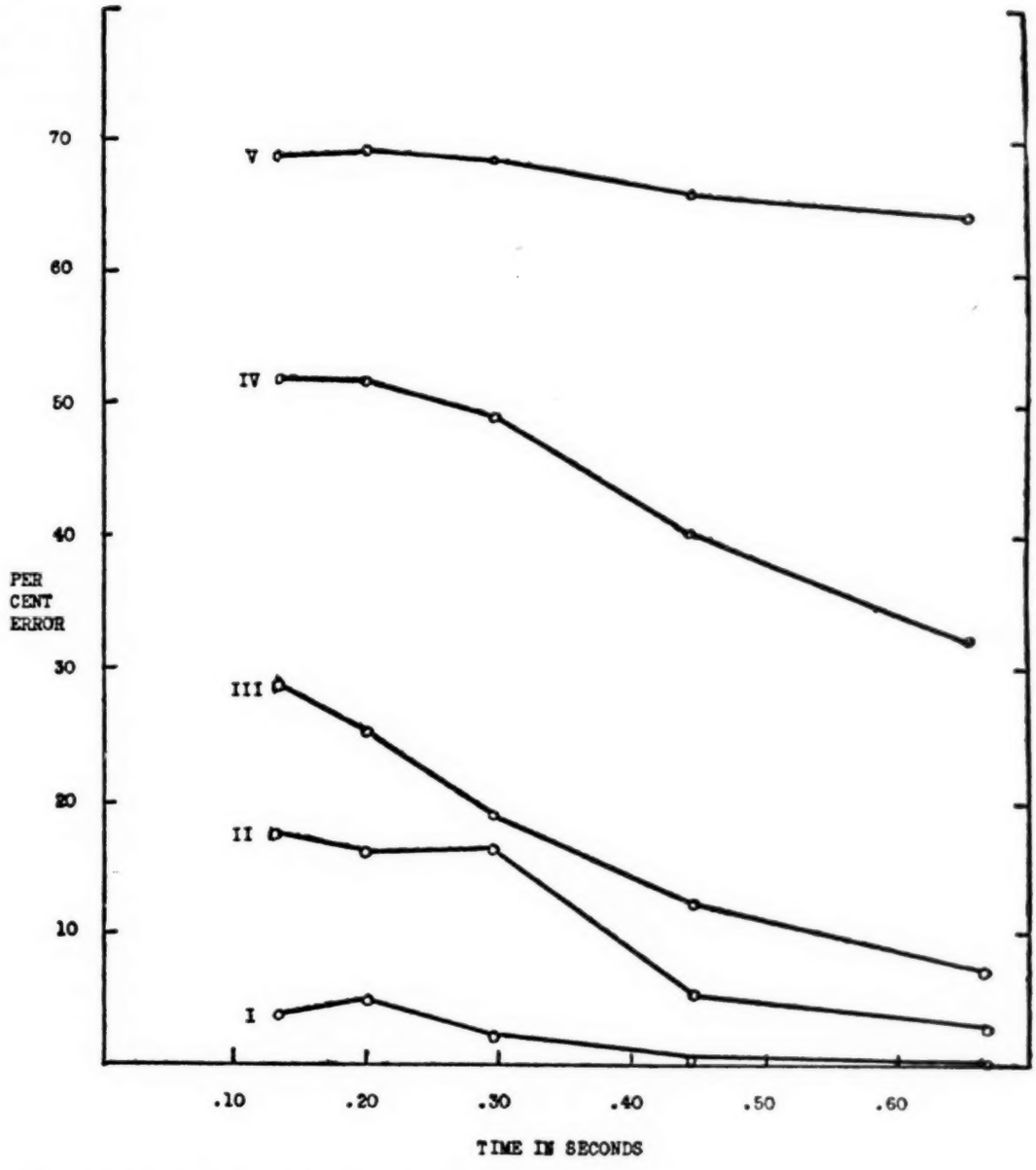

FIG. 2. Relationship between difficulty level and percentage of error. Level I is the easiest level and level V the most difficult.

SECOND EXPERIMENT

In view of the fact that 40 percent of the material of the first experiment was unsuited for the experimental investigation of the problem, since it was too easy or too difficult, a second experiment was planned involving material designed to be more nearly of the same level of difficulty. Furthermore, since the mode of conducting the experiment permitted a chance success of one on four, it was resolved to use a different psychophysical method, *viz.*, the absolute method, where chance success would be less frequent.

The material for this second investigation was designed for color mass discrimination. Cards of the same size as those of the first investigation were used; in this case, however, only two different colors were used at a time and they were randomly distributed according to the following scheme. In a set of 11 cards the same color combination, e.g., blue-green, appeared in varying amounts on all cards. Starting with the largest number of dots of one color, 23 (and hence with 13 dots of the other color), the number of dots of that color diminished progressively by one,

while the number of dots of the other color increased by one. Thus, the ratios of the dots of the two colors in a set of cards were 23-13, 22-14, 21-15,. . . 14-22, and 13-23. These 11 different cards were duplicated to make a set of 22 cards of one color combination, arranged in a randomized order within the set. Six sets, involving all six combinations of the four colors, blue, green, red and purple, two at a time were used; but the blue-red combination, which was the easiest, was used only for practice.

A session consisted of the judgment of the six differently colored sets, 22 cards to a set. The blue-red set, the practice set, was always presented first, and the other sets were presented in a balanced randomized order. Thus, 132 judgments, only 110 of which were recorded, constituted a session.

The series of exposure times used in the previous experiment was modified to yield a wider range of values; the times chosen were .06, .12, .24, .48 and .96 sec. Three sessions were devoted to each of the five exposure times, and the order of their occurrence was randomized from session to session. A schedule of the experimental arrangement, based upon the randomization of cards, sets, and exposure times was precisely followed, and scoring sheets in accordance with these arrangements were made up beforehand to record the responses.

The apparatus and procedure were the same as in the previous experiment. The cards were inserted in the tachistocope, the stimulus was exposed for the predetermined interval, the judgment was made and recorded, and the stimulus card was changed, all operations taking place as rapidly as possible. A brief rest was given at the end of each set.

Three sessions were devoted to each time interval, so that there were 15 sessions, exclusive of the preliminary practice sessions, during which 110 recorded judgments were made, yielding a total of 1650 judgments for each *S*. Seven graduate students, four men and three women, acted as *S*s and each completed the entire series in about five weeks.

The task imposed upon the *S*s differed considerably from that of the previous investigation. When a card was presented the *S* had to rate it on a scale of from 1 to 11, according to the relative amount of designated color on it. Thus, in the blue-green series, the color designated was blue, and the card with the 13 blue dots and the 23 green dots was given the scale rating One; and the card with the 14 blue dots and the 22 green dots, the scale rating Two. After some preliminary practice the *S*s formed their judgments with rapidity and fair confidence.

Scoring was done in two ways: (1) The number of erroneous responses was determined. (b) A weighted number of errors was found by determining the difference between the correct scale value and the scale value given. Thus, if an *S* called a card 'Three' when it should have been called 'Five,' the weighted error for that response was two. For reasons that will be given later, scores by the second method were used to determine the relationship between speed and accuracy in this task.

Results

The total responses of all seven *S*s are given in Table II and in Fig. 3, at every point in the scale, i.e., corresponding to the 11 different

TABLE II

Judgments According to Scale Position

Position	1	2	3	4	5	6	7	8	9	10	11	Aver.
Assigned Judgments	632	755	1156	1302	1337	1174	1218	1256	1004	916	800	1050.0
Correct Judgments	254	186	207	217	185	178	186	203	191	190	293	208.3
Erroneous Judgments	378	569	949	1085	1154	996	1032	1053	813	726	507	841.7
Weighted Errors	2423	2080	1931	1824	1764	1792	1909	1875	2016	2120	2076	1982.7

cards in the set. With the exception of the terminal positions, correct judgments are made with nearly equal frequency at all points of the scale. The sharp increases in correct judgments at the terminal positions are probably due to the fact that those positions serve as frames of reference for the judgments of the series.

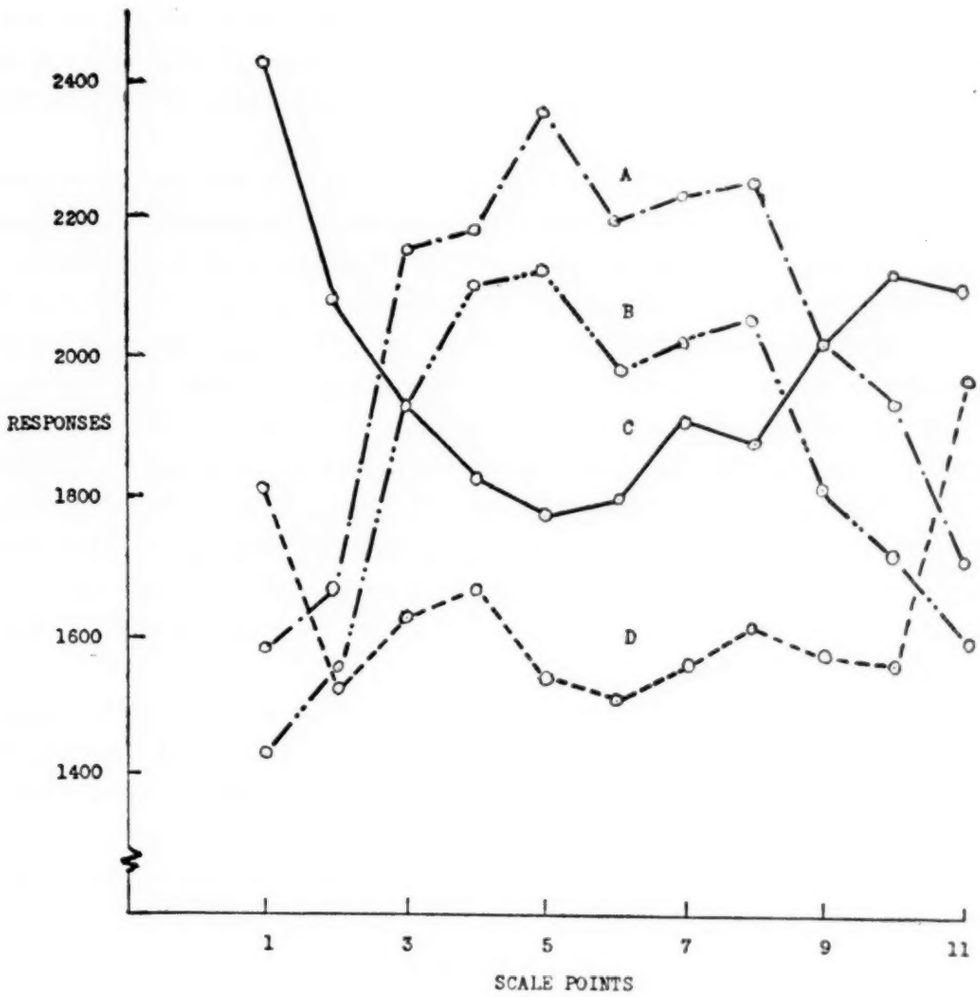

FIG. 3. Responses at each scale point. A—Erroneous responses; B—Assigned responses; C—Weighted errors; D—Correct judgments. The ordinates vary with each curve; those given are for Curve C, the weighted errors.

There is a marked tendency to assign more judgments to the central positions rather than to the terminal positions of the scale, so that a card is rarely called One, Two or Eleven. In other terms, a central tendency effect exists which distorts the results. Since the number of erroneous judgments made at every point of the scale is the difference between the number of assigned judgments and the correct judgments at that point, the curve of erroneous judgments, like that of assigned judgments, shows a pronounced tendency to increase from the terminal to the central positions. Hence, the

number of erroneous judgments varies considerably throughout the scale.

Owing to the central tendency effect, stimuli cards are assigned values which differ more and more from the correct values as one goes from the center of the scale to the extremes. Hence, in spite of the greater precision of judgments at points near the terminal frames of reference, the greatest number of weighted errors is made at the ends of the scale and the least towards the center. Thus it is that the curve of weighted errors is almost the mirror image of that of erroneous responses.

The weighted errors (Method B of scoring) are more evenly distributed over the scale than are the erroneous judgments (Method A), the coefficients of variation being respectively 9 and 30 percent. For this reason the weighted errors were used to decide the form of the relationship between speed and accuracy, although the same curve is obtained for this relationship no matter what scoring method is used. The mode of computing errors yields a variation in the level of difficulty from point to point in the scale, but this variation is considerably less than the discrepancy between the levels of difficulty in the first experiment, and seems to justify our choice of the absolute method, in spite of the complications introduced by the central tendency effect and the tendency for sharper judgments at the ends of the scale.

With the object of determining the practice effect, the weighted errors for complete sets were fractionated in Table III according to the time of exposure of the stimuli. It is to be noted that there is a continuous drop in errors from fastest to slowest speeds. The totals of all three sets are used to give Fig. 1, which is definitely of an ogive type.

TABLE III

NUMBER OF WEIGHTED ERRORS AT DIFFERENT EXPOSURE TIMES AND STAGES OF PRACTICE, TASK II

Sets	Inversions	Exposure Time in Sec.					
		.06	.12	.24	.48	.96	Total
I	6	1706	1741	1604	1342	1239	7632
II	3	1685	1613	1399	1381	1156	7234
III	2	1643	1568	1388	1228	1117	6944
Totals	11	5034	4922	4391	3951	3512	21810

The present data suggest a possible explanation of the discrepancy between Thurstone's hypothesis and Garrett's findings. It will be noted in Table III that only the first set shows an inversion point,

where the number of weighted errors does not decrease with diminishing speed. When the individual records of the *Ss* are studied (Table III, Column 2), it is seen that the number of inversions tends to diminish with practice. The great bulk of the inversions is caused by the fact that the number of errors, for the second fastest exposure time (.12 sec.) is greater than those for the fastest exposure time (.06 sec.). This is not to be wondered at, owing to the slight difference in speed of these two intervals. Probably herein lies the clue to the discrepancy. Optimum points may occur in the initial stages of the experiment, owing to the habitual motor set of the *Ss*; as the experiment progresses this optimum interval may lose its effect. Hence, the continuously decreasing function of speed and accuracy applies only to the judgments of well-practiced *Ss*.

The exposure intervals alone were varied in this experiment and the intervals between exposures were kept approximately constant. Although the cards were always changed as rapidly as possible, the time required to do so was ample for the formulating of a judgment on the stimulus which had previously been presented. It must be emphasized, therefore, that a different function might have been obtained had both exposure time and judgment time been varied.

Summary

This investigation deals with the relationship between exposure time and accuracy, which has been previously shown to follow a sigmoid function for a motor task, in accordance with Thurstone's theory. In this paper it is demonstrated that the sigmoid relationship also holds for two perceptual tasks using different material and studied by different psychophysical methods.

It is to be noted, however, that the sigmoid form may have relatively narrow limits beyond which there is no great variation of accuracy with speed. Till the speed of presentation of the stimuli is increased to the point where discrimination can be made, judgments are given on a chance basis only. Near the upper limits of the function, at the slowest speed, accuracy may decrease very slowly with increasing speed. Obviously, the level of difficulty of material is an important factor in the accuracy of discrimination and will determine the critical limits of the speeds within which the curve shows rapid changes. Since practice tends to diminish the difficulty of the material, a rather precise adjustment of the variables of practice, difficulty level, and speed of presentation is required to show the sigmoid form of the relationship between speed and accuracy.

(Manuscript received March 21, 1946)

References

1. Garrett, H. E. A study of the relation of accuracy to speed. *Arch. Psychol.*, 1922, No. 56.
2. Hunter, W., & Sigler, M. Span of visual discrimination as a function of time and visual stimulation. *J. exp. Psychol.*, 1940, **26**, 160–179.
3. Philip, B. R. The relationship between speed and accuracy in a motor task. *J. exp. Psychol.*, 1936, **19**, 24–50.
4. Philip, B. R. The discrimination of color mass. *Amer. J. Psychol.*, 1938, **51**, 127–135.
5 Philip, B. R. Intra-serial effects in the discrimination of color mass by the ranking method. *J. exp. Psychol.*, 1940, **27**, 125–135.
6. Thurstone, L. L. Ability, motivation and speed. *Psychometrika*, 1937, **2**, 249–254.
7. Young, P. T. *Motivation of behavior*. N. Y.: Wiley, 1936. P. 199.

DEVELOPMENT OF SPEECH DURING INFANCY: CURVE OF PHONEMIC FREQUENCIES *

BY ORVIS C. IRWIN

University of Iowa

In previous articles which have appeared in this journal (1, 3) two problems of the increasing mastery of speech sounds by infants were considered. The first presented equations and curves of the development of phonemic types. The second was concerned with the Chen Differential Percentage Index. Other studies have dealt with the problems of vowel and consonant types (2) and vowel and consonant frequencies (4). The purposes of this report is to give an analysis of phoneme frequency during the first two and a half years of life, to indicate equations derived from the data, and to present curves of phoneme frequency development representative of the data.

Three analyses of the data have been made. The first considers records obtained upon a group of 35 infants which was followed continuously throughout the period of infancy, the second treats the data collected on 95 Ss during this period, and the third is concerned with the problem of sex differences.

The data for the group yielding the most continuous records are given in Table I, which includes bimonthly means of the total frefrequency of speech sounds, the number of infants on which records were taken, and the number of records. Each record consists of such sounds as occurred on 30 breaths. The table should be read as follows: for the first two month period the number of infants studied was 25, the number of records taken was 52, the mean frequency was 61.8, and the standard deviation 9.3.

In order to secure close fitting curves the data for the entire $2\frac{1}{2}$-year period were divided into two parts; one part included the means for the first $1\frac{1}{2}$ years, the other for the remaining year. Separate equations were derived for each. The curves are shown in Fig. 1. The equation for the means of the first $1\frac{1}{2}$ years is $P_F = 4.525\,A + 58.0$, in which P_F is the phoneme frequency and A is age. The equation for the second part of the period is $P_F = 31\,e^{.1673\,A} + 56.3$. The first equation is represented by a straight line, the second by an exponential curve. It is obvious by inspection that the theoretical curve is a reasonably good fit to the empirical data. The graph

* The term 'frequency' refers to the number of times phoneme types occur. 'Type' refers to the speech sounds listed in the International Phonetic Alphabet.

TABLE I

MEANS OF PHONEME FREQUENCIES
(Most Continuous Data)

Bimonthly Age Group	Number of:		Means	SD
	Subjects	Records		
1–2	25	52	61.8	9.3
3–4	32	74	66.6	10.9
5–6	32	71	73.1	10.0
7–8	33	94	75.8	13.2
9–10	34	87	80.4	14.4
11–12	35	91	90.4	19.1
13–14	34	81	89.0	20.7
15–16	34	85	91.6	15.1
17–18	34	74	96.8	16.7
19–20	31	70	95.5	15.8
21–22	31	62	103.9	23.2
23–24	27	48	113.3	29.7
25–26	28	47	133.2	34.8
27–28	24	41	140.3	40.7
29–30	19	27	155.8	42.2

indicates that for the first 1½ years the average infant's production of phoneme frequencies increases at a constant rate and that thereafter the rate is an increasing one.

A test of the validity of this curve would be to secure more cases and to combine them in order to discover whether the curve is thereby modified. This has been done. The data for all infants on whom records were taken throughout the period of infancy are found in the next table.

TABLE II

MEANS FOR PHONEME FREQUENCIES
(All Data)

Bimonthly Age Group	Number of:		Means	SD
	Subjects	Records		
1–2	62	125	60.3	9.1
3–4	80	181	67.1	10.9
5–6	75	166	73.9	11.7
7–8	64	170	76.4	14.1
9–10	62	147	81.1	18.9
11–12	62	149	91.5	19.2
13–14	57	127	88.2	20.1
15–16	55	128	92.4	16.6
17–18	50	106	97.8	17.2
19–20	41	84	98.1	20.9
21–22	37	68	107.1	28.1
23–24	31	52	115.1	28.4
25–26	32	51	137.8	38.0
27–28	24	41	140.3	40.7
29–30	19	27	155.8	42.2

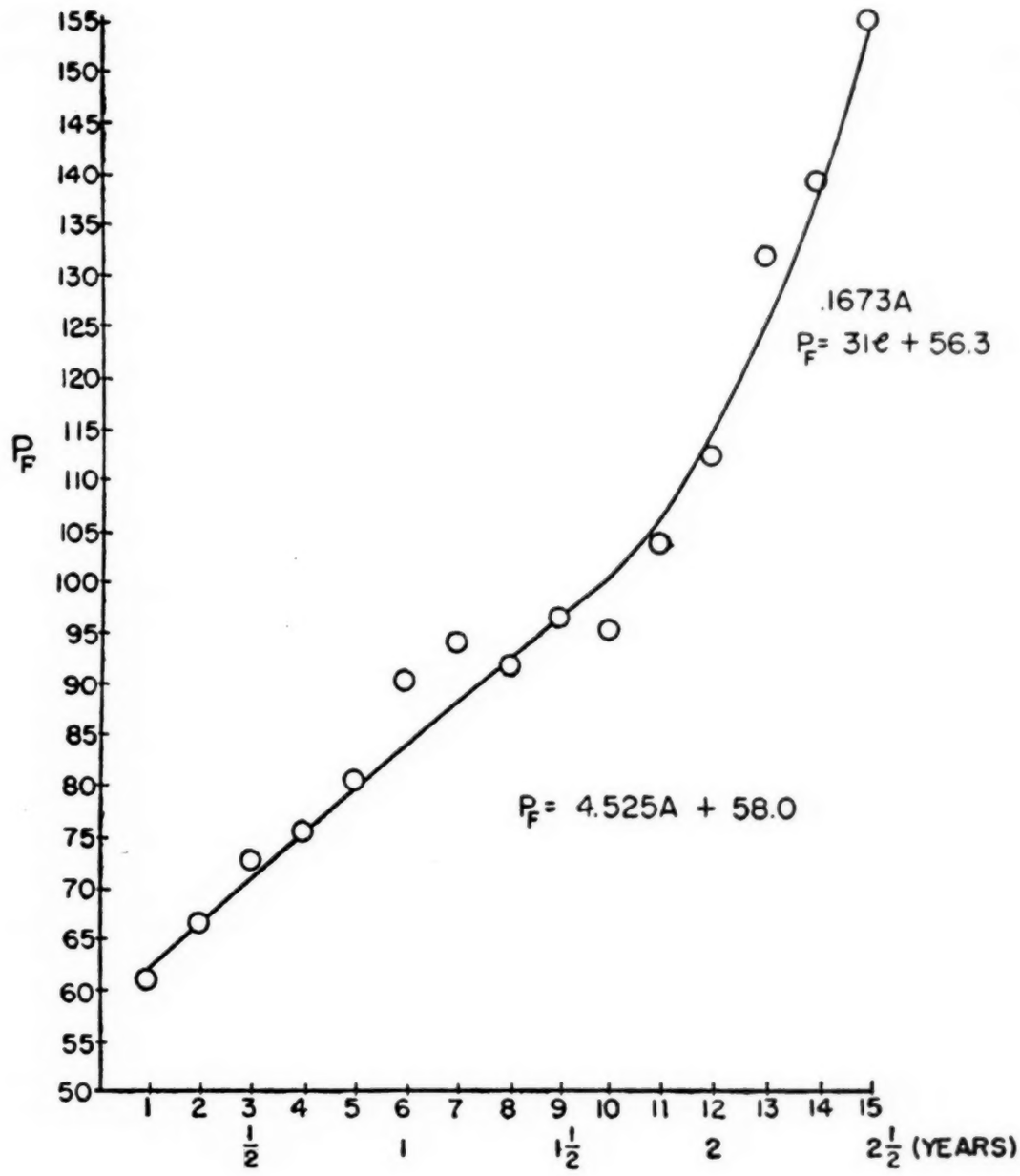

FIG. 1. Phoneme frequency (most continuous data)

Again it has been necessary to divide the data into two parts and derive separate equations. The curves appear in Fig. 2. The equation for the first $1\frac{1}{2}$ years is $P_F = 4.493\ A + 58.5$, and for the second part it is $P_F = 27\ e^{.1774\ A} + 63.4$.

It is seen that the constants in the equations for the straight lines fitted to the first parts of the two sets of data are practically identical. Moreover, the values of the exponents in the equations for the last parts of the data are nearly the same. Thus the addition of more data does not change the rate as given by the first curve. The curves in both graphs indicate that infants produce about 60 phoneme frequencies in the first bimonthly period and 156 in the last period. It

therefore is reasonable to conclude that the increase in phoneme frequency proceeds for the first 1½ years at a constant rate and that for the remaining period it follows an exponential function.

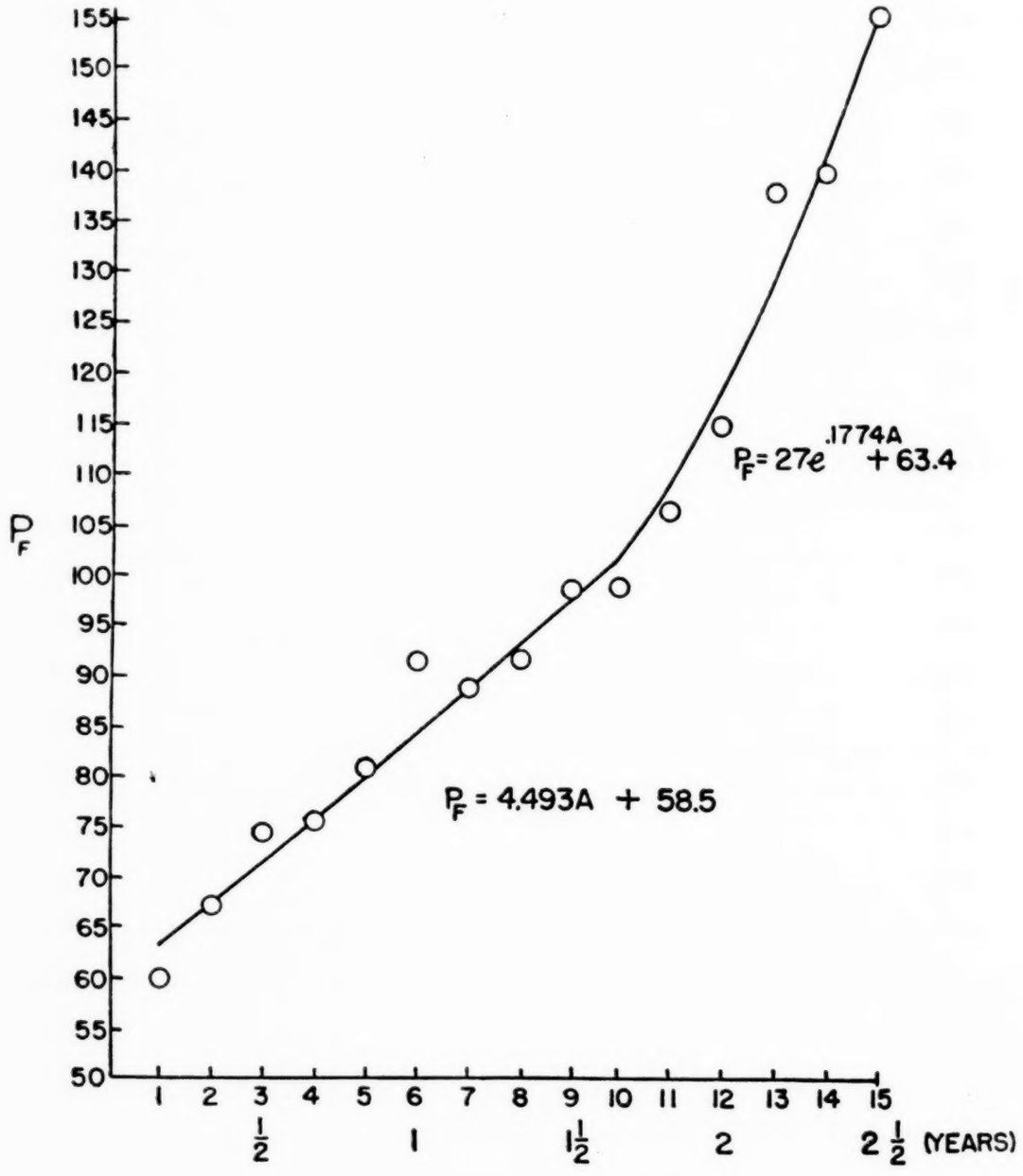

FIG. 2. Phoneme frequency (all data)

When these curves for the progress of phoneme frequency are compared with those for phoneme type (1), an interesting contrast becomes apparent. The type curves are parabolic in nature, whereas the frequency curves are at first linear and then exponential. In terms of these two criteria, this means that there are fundamental rate differences in speech development during infancy. The rate of acquisition of phoneme types exhibits a decelerating increase, but for

phoneme frequency the increase is at first constant and finally becomes an accelerated rate.

A last analysis is concerned with the problem of sex differences in infant production of sounds. The data are given in Table III.

TABLE III

MEANS OF PHONEME FREQUENCIES
SEX DIFFERENCES

Bimonthly Age Group	Boys				Girls			
	Number of:		Means	SD	Number of:		Means	SD
	Subjects	Records			Subjects	Records		
1–2	13	27	64.4	9.5	12	25	59.1	8.1
3–4	16	35	71.3	12.1	16	39	62.0	7.4
5–6	16	37	74.7	9.4	16	34	71.5	10.4
7–8	17	53	79.9	13.1	16	41	71.5	11.6
9–10	18	48	80.4	17.0	16	39	80.4	10.9
11–12	18	52	82.4	15.2	17	39	98.9	19.0
13–14	18	47	86.3	18.9	16	34	91.9	22.4
15–16	17	46	88.0	14.6	17	39	95.3	14.6
17–18	17	42	97.0	15.9	17	32	96.7	17.5
19–20	16	37	91.0	14.8	15	33	100.2	15.4
21–22	17	35	103.3	25.9	14	27	104.6	19.0
23–24	16	29	116.3	32.5	11	19	109.0	24.8
25–26	16	32	137.7	37.1	12	15	127.1	30.8
27–28	13	20	134.5	46.3	11	21	147.2	30.9
29–30	10	15	162.6	47.2	9	12	148.3	33.7

It was found that in regard to sex differences exponential equations yielded closely fitting curves. The equations are: $P_F = 10\ e^{.1585\ A} + 57.5$ for the means of boys, and $P_F = 25\ e^{.0947\ A} + 38.0$ for girls. The curves are plotted in Fig. 3.

An examination of these curves shows that girls exhibit a slightly lower frequency than boys, but that they develop for $1\frac{1}{2}$ years at a rate slightly in excess of that for boys. Thereafter the boys' rate exceeds that of girls. The differences however are not statistically dependable.

When the curves of sex differences in phoneme types (1) are compared with those for phoneme frequency, two interesting reversals are apparent. The curves for phoneme types are parabolic in nature, while those for frequency are exponential. Thus the rate for the former is a decelerating one during the period of infancy, while the latter is an accelerating one. Moreover, in the case of the development of phoneme types there seems to be a final tendency for girls to exceed boys, whereas in regard to frequency of sound production there is an opposite tendency.

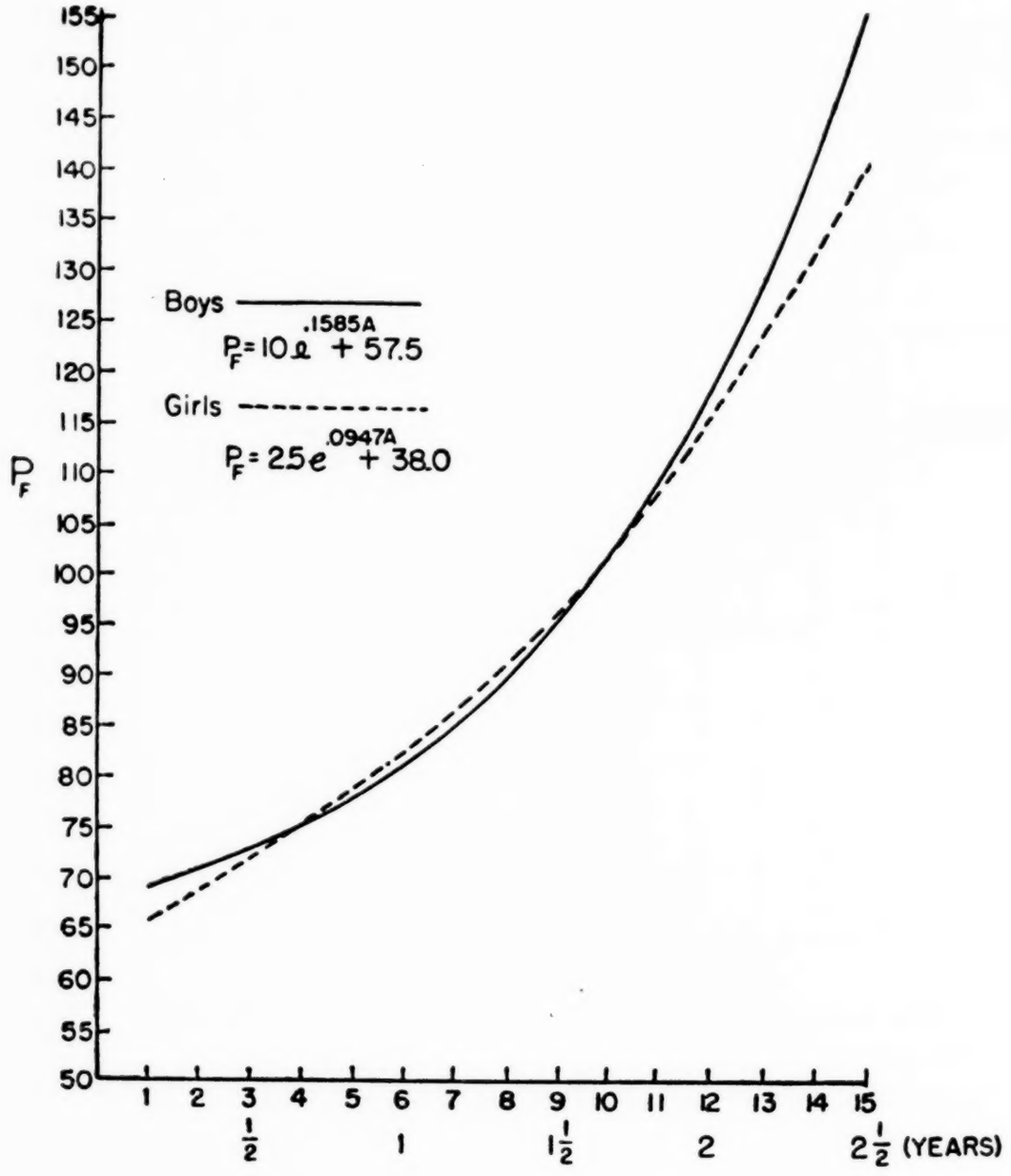

FIG. 3. Phoneme frequency sex differences

SUMMARY

This article reports equations and curves derived from data on the frequency of sound production of infants under $2\frac{1}{2}$ years of age. A curve of frequency of vocalization of the 35 most continuously observed children among a group of 95 is given. Separate equations were derived for the data. The one covering the first year and a half is $P_F = 4.525\,A + 58.0$, the other covering the remaining year is $P_F = 31\,e^{.1673\,A} + 56.3$. This means that the rate of frequency development is at first constant and then becomes exponential. An equation derived from the data of all the Ss yields a similar curve. When these curves are compared with those for phoneme types pre-

viously reported, the latter are seen to be parabolic, whereas the former are at first linear and then exponential. In regard to sex differences, exponential equations yield the best-fitting curves. While boys have a tendency to exceed girls in frequency of sound production during the last six month of the period, the differences are not statistically significant.

(Manuscript received March 28, 1946)

References

1. Chen, H. P., & Irwin, O. C. Development of speech during infancy: Curve of differential percentage indices. *J. exp. Psychol.*, 1946, **36**, 522–525.
2. Chen, H. P., & Irwin, O. C. Infant speech: vowel and consonant types. *J. Speech Disorders*, 1946, **11**, 27–29.
3. Irwin, O. C., & Chen, H. P. Development of speech during infancy: Curve of phonemic types. *J. exp. Psychol.*, 1946, **36**, 431–436.
4. Irwin, O. C., & Chen, H. P. Infant speech: vowel and consonant frequency. *J. Speech Disorders*, 1946, **11**, 123–125.

THE MEANING OF KOCH AND DANIEL'S RESULTS FOR HULL'S THEORY

BY CHARLES B. WOODBURY

In any attempt to utilize a theoretical system in explaining or predicting experimental results, it is logically necessary that *all* of the relevant theorems of the system be used. Conversely, when empirical findings are used as a basis for modifying such a system, it is important to examine the system as a whole, in order to determine just what modification is required. Generally in scientific practice it is deemed preferable to make necessary revisions in such a way that the fundamental postulates of the system are changed as little as possible.

The purpose of this note is to show how Koch and Daniel's results (2) can be reconciled with Hull's theory without forcing any change in the present formulation of its postulates.

Their finding, that rats trained in a Skinner-box habit under strong motivation showed practically zero response strength when tested immediately after satiation, conflicts with the extrapolation of the curve which Perin found for response strength as a function of drive strength (3). This curve indicates that even at zero drive this habit would show from 28 percent to 50 percent of its maximum strength, depending on the measure used. The closest approach which Perin's experiment made to this, however, was a group of animals run one hour after satiation.

As Koch and Daniel point out, their finding necessitates a change in the form of Perin's curve for reaction strength as a function of hunger so that, starting practically at zero, the curve would rise rapidly to a point at about one to three hours of hunger, and thereafter rise with the slight positive acceleration which Perin found.

It is not necessary, however, as these authors believe (2, p. 170), to revise Hull's theoretical formulation of the dependence of excitatory potential on the variables of habit strength and drive strength, if another of his postulates is included in our thinking. This is Postulate II, which states,

> All afferent neural impulses (s) active in the nervous system at any given instant, interact with each other in such a way as to change each into something partially different ($\check{s}$) in a manner which varies with every concurrent associated afferent impulse or combination of such impulses. Other things equal, the magnitude of the interaction effect of one afferent impulse upon a second is an increasing monotonic function of the magnitude of the first (1, p. 47).

Now, if Hull's definition of excitatory potential be examined (1, p. 404), it appears that it is a function of *effective* habit strength, which in turn is a function (a) of the number of times the habit has been reinforced, and (b) of the distance, on the generalization gradient, between the stimulus at the

time of measurement and the stimulus present during the original training. Furthermore, in Hull's system the stimulus aspect of drive is explicitly recognized,[1] and drive stimuli are held to be subject to weakening along a generalization gradient just as others are (1, p. 235).

In the present experiments, therefore, it is clear that the change in drive conditions during extinction changes the stimulus pattern from that to which the reaction was conditioned. This, by the principle of generalization (1, pp. 216–217), must reduce the effective habit strength at the time of extinction, independently of any purely motivational effect of drive strength on excitatory potential.[2]

To clarify this point it may be well to indicate some of the ways in which stimulation was different during extinction from what it had been in the training trials. The only difference in procedure, of course, was the satiation of the rats just before the extinction tests, but this alone changes the stimulating conditions in at least two ways: (a) the internal stimuli arising from food deprivation (stomach contractions, salivation, etc.) were absent during extinction; and (b) stimuli arising from satiation (stomach distention, etc.) were present during extinction but had been absent in training. In addition, all but the first of the training trials had been preceded by reinforced responses, a condition not present during extinction; and, in the majority of training trials, the rat had been in the box for some minutes at the time of reinforcement, whereas resistance to extinction was necessarily tested immediately upon placing the rat in the box.

Because of these changes in stimulation under the conditions for extinction, it is clear that the *effective* habit strength would be expected to drop, and we can reason that the zero excitatory potential is due to the drop of effective habit strength, rather than to the existence of a different relationship than that which Hull postulates between drive strength and excitatory potential. It is, therefore, possible that if the effects of stimulus change could be partialled out, the habit might give rise to an appreciable excitatory potential even at zero drive strength. In this case Perin's equation may fairly adequately represent the relation between drive strength and excitatory potential, assuming effective habit strength to remain constant (which of course it cannot under the present experimental conditions). The correction which Koch and Daniel's data impel would enter into the calculation of effective habit strength (of which excitatory potential is a function), rather than directly into the dependence of excitatory potential on drive strength. Similarly, the corollaries which depend on Postulate VII (which expresses this relationship), and Hull's assumption concerning the role and strength of 'residual' drives, cannot be unequivocally ruled out on the basis of the present data alone, although certainly Koch and Daniel's results seem to cast doubt on them.

[1] "The afferent discharges arising from the drive stimulus (S_D) become conditioned to reactions just the same as any other elements in stimulus compounds . . ." (1, p. 240).

[2] The motivational aspect of drive is not, in Hull's system, a determinant of the strength of habit as such; the habit remains the same, though drive may be stronger or weaker. Rather, the strength of habit *and* the strength of drive at a given time determine the excitatory potential, which in turn determines the strength of reaction.

There still remains the problem of why Perin's study failed to show excitatory potential to approach zero as drive was decreased and as, presumably, effective habit strength approached zero. I believe that this is because the modification of stimulation which results from reducing drive is very much greater in the first hour, probably the first half hour, after satiation than it is later; that is, the internal stimuli probably change most rapidly just after eating. This would result in the rapid rise in effective habit strength, and hence of excitatory potential, which Koch and Daniel indicate must take place in the first hour or so after satiation. By this hypothesis, Hull's assumption of a linear relationship between drive stimuli and drive strength (1, p. 240) appears to be incorrect; in the light of Koch and Daniel's results it would appear that drive stimuli change quite sharply as drive strength approaches very close to zero.

The present analysis shows how Koch and Daniel's results can be reconciled with Hull's formal system without requiring a change in any of its basic postulates. At the same time, it is obvious that before the truth of either this or Koch and Daniel's interpretation can be asserted, careful studies of the relationship between the strength of drive and the characteristics of the resulting stimulation, and of the exact role of drive stimuli in learning, will be necessary.

(Manuscript received April 22, 1946)

References

1. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
2. Koch, S., & Daniel, W. J. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1945, **35**, 167–187.
3. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.